



---

# **TECHNICKÁ UNIVERZITA V LIBERCI**

Fakulta mechatroniky, informatiky a mezioborových studií

Studijní program: N2612 – Elektrotechnika a Informatika

Studijní obor: 1802T007 – Informační technologie

## **Objektivní vyhodnocování kvality řečových signálů**

## **Objective quality assessment of speech signals**

### **Diplomová práce**

Autor:	<b>Bc. Jan Pavlík</b>
Vedoucí práce:	Ing. Zbyněk Koldovský, Ph.D.
Konzultant:	Ing. Jiří Málek, Ph.D.

**V Liberci 17. 5. 2012**



## **Prohlášení**

Byl(a) jsem seznámen(a) s tím, že na mou diplomovou práci se plně vztahuje zákon č. 121/2000 Sb., o právu autorském, zejména § 60 – školní dílo.

Beru na vědomí, že Technická univerzita v Liberci (TUL) nezasahuje do mých autorských práv užitím mé diplomové práce pro vnitřní potřebu TUL.

Užiji-li diplomovou práci nebo poskytnu-li licenci k jejímu využití, jsem si vědom povinnosti informovat o této skutečnosti TUL; v tomto případě má TUL právo ode mne požadovat úhradu nákladů, které vynaložila na vytvoření díla, až do jejich skutečné výše.

Diplomovou práci jsem vypracoval(a) samostatně s použitím uvedené literatury a na základě konzultací s vedoucím diplomové práce a konzultantem.

Datum

Podpis

## **Poděkování**

Tímto bych chtěl poděkovat panu Ing. Zbyňkovi Koldovskému, Ph.D. za zadání a vedení mé diplomové práce. Dále bych chtěl poděkovat všem účastníkům poslechových testů, bez kterých by tato práce nešla dokončit, a kteří mě ochotně věnovali část svého času. Mé díky patří i mluvčím, kteří namluvili řečové vzorky. V neposlední řadě chci vyjádřit vděčnost i mé rodině a přátelům, kteří mě po celou dobu tvorby práce podporovali.

## Abstrakt

Diplomová práce se zabývá vyhodnocováním kvality řečových signálů a to pomocí několika metod. Zástupcem subjektivních metod jsou poslechové testy, jako objektivní metoda je pak použita metoda PEMO-Q. Kvalita řečového signálu je ovlivněna i jeho hlasitostí. K vyhodnocení hlasitosti je použit váhový filtr A a equal-loudness filtr. V první části práce je popsán řečový signál a je vysvětleno, co se skrývá pod pojmem kvalita řečového signálu. V další části naleznete popis poslechových testů, metody PEMO-Q a aparátů na vyhodnocení hlasitosti signálů. V praktické části diplomové práce se seznámíte s postupem vytvoření databáze testovacích signálů a s dosaženými výsledky. Výstupem práce je rozbor databáze řečových signálů s aditivním šumem a porovnání poslechových testů a metody PEMO-Q.

**Klíčová slova:** vyhodnocení kvality, řečový signál, poslechové testy, PEMO-Q, hlasitost

## Abstract

This thesis deals with the quality assessment of speech signals using several methods. The representatives of subjective methods are listening tests, while as an objective method the PEMO-Q method is used. The quality of speech signal is influenced by its volume. To evaluate loudness, weighting filter A and equal-loudness filter are used. In the first section of the thesis, it is described speech signal and it is explained what is meant by the quality of the speech signal. The next section describes the listening tests, the PEMO-Q method and the apparatus for evaluating the loudness of signals. The practical part of thesis presents how a database of test signals was created. It also presents achieved results. The output of the work is the analysis of database of speech signals with additive noise and a comparison of the listening test with the PEMO-Q method.

**Key words:** quality evaluation, speech signal, listening tests, PEMO-Q, loudness

# Obsah

Prohlášení.....	3
Poděkování.....	4
Abstrakt.....	5
Abstract.....	5
Obsah .....	6
Seznam ilustrací.....	8
1 Úvod .....	10
2 Řečový signál .....	11
2.1 Kvalita řečového signálu.....	12
3 Hodnocení kvality řečových signálů .....	15
3.1 Subjektivní metody (Poslechové testy).....	15
3.1.1 ITU-R BS.1116.....	16
3.1.2 ITU-T P.835.....	17
3.2 Metody pro objektivní vnímání kvality zvuku.....	19
3.2.1 PEMO-Q.....	20
3.2.1.1 Charakteristika metody.....	20
3.2.1.2 Popis metody .....	21
4 Hlasitost signálu .....	25
4.1 Váhový filtr A .....	25
4.2 Křivky konstantní hlasitosti (Equal-loudness).....	26
5 Databáze řečových signálů .....	28
6 Vyhodnocení poslechových testů .....	30
6.1 Šum v pozadí signálu .....	30
6.2 Kvalita řečového signálu.....	32
6.3 Celková kvalita signálu .....	34

7	Vyhodnocení objektivní metody PEMO-Q .....	36
7.1	Objektivní kvalita řeči $q_c$ .....	36
7.2	Měření vnímané podobnosti PSM .....	38
7.3	Srovnání poslechových testů a objektivních metod .....	43
8	Vyhodnocení hlasitosti řečových signálů .....	44
8.1	Čisté řečové vzorky .....	44
8.2	Řečové vzorky s aditivním šumem .....	45
8.3	Zfiltrované řečové vzorky s aditivním šumem .....	48
9	Závěr .....	50
	Seznam použité literatury .....	52
	Příloha A – Instrukce k poslechovému testu .....	54



## Seznam ilustrací

Obr. 2.1: Řečový signál

Obr. 2.2: Řečový signál - znělé

Obr. 2.3: Řečový signál - neznělé

Obr. 2.4: Vlivy ovlivňující kvalitu řečového signálu

Obr. 2.5: Znázornění oblasti kvality přeneseného hlasu

Tab. 3.1: Pětistupňová stupnice pro hodnocení kvality zvukového signálu

Tab. 3.2: Kvalita řečového signálu

Tab. 3.3: Rušení pozadí

Tab. 3.4: Celková kvalita signálu

Obr. 3.5: Blokové schéma poslechového modelu

Obr. 3.6: Blokové schéma metody PEMO-Q

Tab. 4.1: Tabulka korekcí váhového filtru A

Obr. 4.2: Equal-loudness křivky

Graf 6.1: Šum v pozadí

Tab. 6.2: Porovnání stacionárních a nestacionárních šumů

Graf 6.3: Řečový signál

Graf 6.4: Celková kvalita signálu

Tab. 6.5: Porovnání stacionárních a nestacionárních šumů

Graf 7.1: Objektivní kvalita řeči  $q_c$

Graf 7.2: Celková kvalita signálu

Tab. 7.3: Porovnání stacionárních a nestacionárních šumů

Graf 7.4: PSM

Graf 7.5: Celková kvalita signálu

Tab. 7.6: Porovnání stacionárních a nestacionárních šumů

Graf 7.7: PSMt

Graf 7.8: Hovorový kanál

Obr. 8.1: Váhový filtr a equal-loudness filtr

Tab. 8.2: Vliv mluvčího na hlasitost

Graf 8.3: Změna SNR po aplikaci váhového filtru

Graf 8.4: Šum v pozadí

Graf 8.5: Změna SNR po aplikaci váhového filtru

Graf 8.6: Změna SNR po aplikaci filtrů – muž

Graf 8.7: Změna SNR po aplikaci filtrů – žena

# 1 Úvod

Komunikace mezi lidmi na dlouhé vzdálenosti je běžná záležitost a v současné době i nutnost. Mezi nejpoužívanější druh komunikace patří telekomunikace, čili přenos hlasu. V oblasti telekomunikace je snaha vyvíjet stále nové technologie tak, aby přenos hlasu byl co nejspolehlivější a co nejvíce kvalitní. Na druhou stranu je kladen důraz na co nejmenší datovou náročnost přenesené řeči. Kvalita přenosu hlasu je jedním z měřítek k porovnání různých telekomunikačních technologií. Proto vyhodnocování kvality řečových signálů hraje důležitou roli při vývoji současných telekomunikačních sítí.

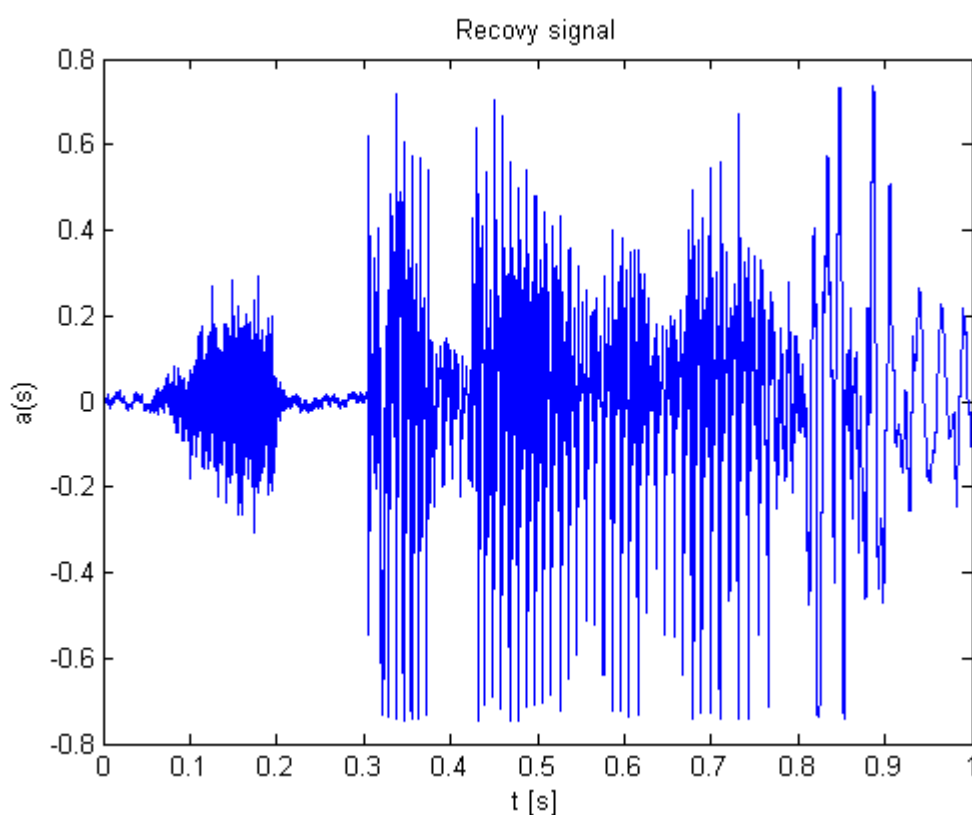
Ke snížení kvality řečového signálu může dojít mnoha způsoby, ať už nějakým rušivým šumem na straně mluvčího, nebo zkreslením signálu při přenosu. Nejběžnějším způsobem jak vyhodnotit kvalitu řečového signálu jsou poslechové testy, kterých existuje více variant pro různé typy rušení. Největší výhodou poslechových testů je to, že při správném provedení je můžeme brát jako absolutní měřítko pro posouzení kvality řečového signálu. Mají však i dost nevýhod, například časovou náročnost. Proto existují snahy o vytvoření metod, při kterých kvalitu nehodnotí člověk, jak je tomu u poslechových testů, ale vyhodnocení probíhá automaticky nějakým počítačovým programem, který se snaží simulovat poslechové testy.

V této práci tedy provedu poslechové testy na uměle vytvořené databázi řečových signálů, u kterých je kvalita snížena aditivním šumem i zkreslením způsobeným průchodem filtrem. Stejně testovací signály pak vyhodnotím pomocí nové metody pro objektivní vyhodnocení vnímané kvality PEMO-Q (Perceptual Model-Quality assesment). Dosažené výsledky porovnam.

Jedním z aspektů ovlivňující kvalitu zvuku je jeho hlasitost. Proto se budu zabývat tím, jak aditivní šum nebo průchod filtrem ovlivní hlasitost signálu. K tomu účelu využiji aparáty, které berou v potaz citlivost lidského ucha v závislosti na frekvenci zvuku, čili váhový filtr A a equal-loudness filtr.

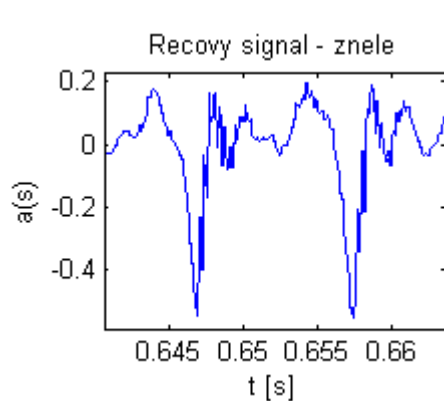
## 2 Řečový signál

Lidská řeč je nositelem užitečné informace, která je přenášena pomocí akustického vlnění mezi řečníkem a posluchačem. Je to vlastně souvislý, časově proměnný proces. Lidská řeč je vytvářena vydechováním vzduchu z plic přes hlasová ústrojí člověka (hlasivky, jazyk, rty). Převědeme-li řeč například pomocí mikrofону na odpovídající řečový signál, vznikne nám posloupnost diskrétních vzorků. Řečový signál vzniklý vyslovením slova UČEBNA si můžeme prohlédnout na obrázku (Obr. 2.1: Řečový signál).

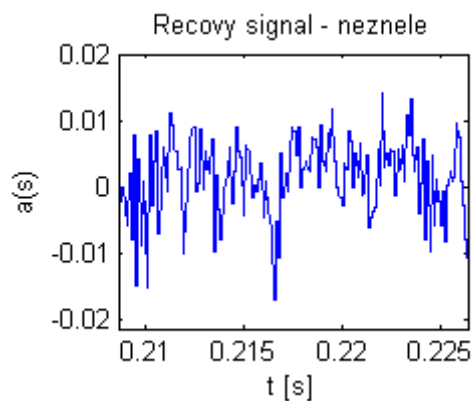


Obr. 2.1: Řečový signál

Rozebereme-li si tento signál, najdeme v něm dvě odlišné oblasti. První oblast obsahuje tzv. znělé části řeči (Obr. 2.2: Řečový signál - znělé), tyto části jsou víceméně periodické. Oproti tomu se zde vyskytují části, které mají charakter šumu a nazývají se neznělé (Obr. 2.3: Řečový signál - neznělé).



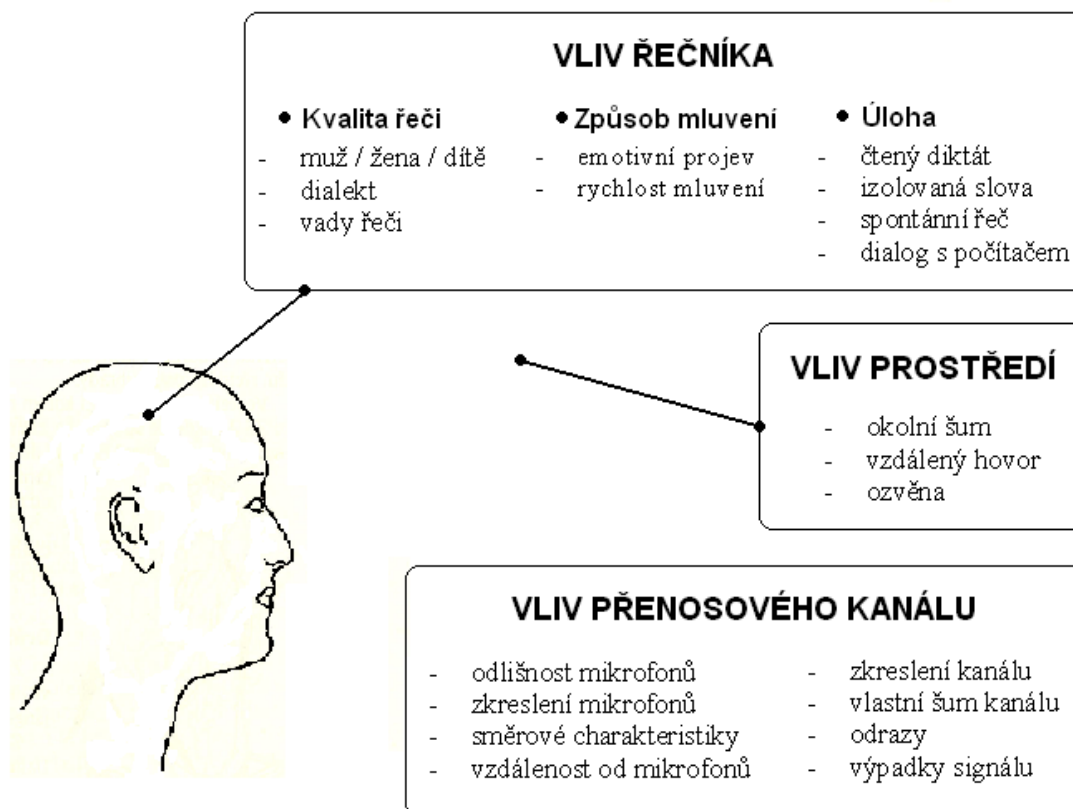
Obr. 2.2: Řečový signál - znělé



Obr. 2.3: Řečový signál - neznělé

## 2.1 Kvalita řečového signálu

Pod pojmem kvalita řečového signálu si můžeme představit mnoho věcí. Jedna z definic může znít takto: Kvalita řečového signálu je subjektivní hodnocení jakosti řečového signálu přeneseného komunikačním kanálem. Kvalita nám vlastně popisuje znehodnocení řečového signálu způsobené souhrnem několika aspektů, které jsou znázorněny na obrázku (Obr. 2.4: Vlivy ovlivňující kvalitu řečového signálu). Dalším výkladem kvality zvukového signálu může být míra podobnosti signálu A vyslaného z jedné stanice a signálu A' přijatého druhou stanicí.



Obr. 2.4: Vlivy ovlivňující kvalitu řečového signálu [6]

Jak bylo řečeno v [6], mezi aspekty, které významně ovlivňují kvalitu řečového signálu, patří:

- **Srozumitelnost**

Srozumitelnost je přímo vázaná na vnímání hlasu lidským uchem a chápeme ji jako věrnost a zřetelnost původního signálu. Může být také chápána jako množství informace, kterou jsme schopni ze zprávy získat. Na čistotu zvuku mají vliv převážně působení dočasných výpadků, ztráty signálu, šumy, zesílení a přenosové chyby kanálu.

- **Zpoždění end-to-end**

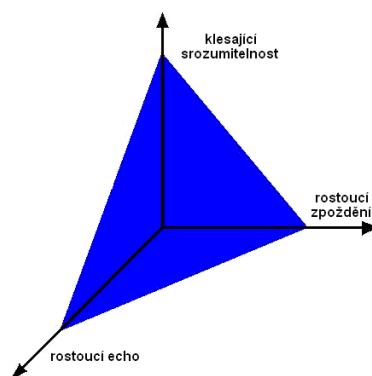
Tvoří součet všech zpoždění, která působí na přenášený signál po celé přenosové cestě. Zpoždění neovlivňuje kvalitu řeči přímo, ale ovlivňuje kvalitu konverzace. Zpoždění do 100ms nejsou postřehnutelná, při zpoždění

v rozmezí 100 – 300ms je patrná prodleva v odpovědích a u zpoždění nad 300ms se komunikace stává nemožnou.

- **Echo**

Jedná se o zvuk, který se vrací zpět ke zdroji. Pokud je mezi původním a odraženým signálem (echem) malý časový rozdíl (do 30ms), je vliv echa zanedbatelný. Při větším časovém rozdílu je jeho existence nežádoucí. Echo vzniká například akustickou vazbou mezi reproduktorem a mikrofonem, zakončovacími obvody v síti PSTN (Public Switched Telephone network) apod.

Na obrázku si můžeme prohlédnout znázornění oblasti kvality přeneseného hlasu, čím blíže se pohybujeme k průsečíku os, tím je kvality přenosu hlasu lepší.



Obr. 2.5: Znázornění oblasti kvality přeneseného hlasu [6]

Další parametry ovlivňující kvalitu řeči jsou např. hlasitost – subjektivní pocit, kterým hodnotíme intenzitu akustického vjemu, přirozenost – stupeň věrnosti hlasu mluvčího, šum, přeslechy a další.

### 3 Hodnocení kvality řečových signálů

Hodnocení kvality řečových signálů hraje klíčovou úlohu v telekomunikační technice. V tomto oboru vznikají stále nové technologie, nové způsoby zpracování řeči, roste komplikovanost a robustnost telekomunikačních sítí. Je zde požadavek na snižování velikosti přenesených dat při zachování co největší kvality řečového signálu. Kvalita přeneseného hlasu je tudíž jedno ze základních měřítek pro porovnání nových technologií či metod pro zpracování zvuku.

Chceme-li vyhodnotit kvalitu řečových signálů, můžeme zvolit jeden z těchto dvou základních přístupů:

- **Subjektivní metody**
- **Objektivní metody**

#### 3.1 Subjektivní metody (*Poslechové testy*)

V dnešní době existuje spousta metod, jak co neoptimálněji vyhodnocovat kvalitu řečového signálu. Největší důraz je kladen na takové metody, které posuzují kvalitu z hlediska lidského vnímání. Lidské vnímání zvuku je však velice subjektivní, záleží na věku či pohlaví posluchače, vyvinutosti sluchu a mnoho jiném. Popis subjektivních metod můžeme nalézt v doporučeních organizace ITU (International Telecommunication Union).

Klasickým způsobem stanovení kvality zvukových signálů jsou subjektivní poslechové testy. Při správném provedení nám tyto testy dávají nejvíce vypovídající výsledky. Na druhou stranu se s těmito testy pojí i značné nevýhody, mezi které můžeme zařadit zejména náročnost v přípravě, provedení i vyhodnocení testů. Prvním krokem k úspěšnému provedení poslechového testu je určit si jasný cíl, tedy co přesně chceme poslechovým testem zjistit, a tomu uzpůsobit i způsob provedení testu. Výsledek testu ovlivňuje návrh sady testových vzorků, výběr správného reprezentativního vzorku respondentů, korektní statistické vyhodnocení výsledků. Je zde mnoho aspektů, kde může dojít k chybě, proto je u poslechového testu velká možnost, že nedostaneme odpověď, kterou jsme hledali. V neposlední řadě je



při provádění poslechových testů kladen velký důraz na poslechové podmínky, tj. akusticky vyhovující prostor, adekvátní reprodukční zařízení, vhodná konfigurace. Ze všeho výše uvedeného vyplývá, že poslechové testy jsou velmi časově i finančně náročné.

I přes tyto nevýhody jsou poslechové testy stále jediným absolutním měřítkem vnímané kvality řečového (i jakéhokoli jiného zvukového) signálu, neboť jinými metodami vždy dostaneme výsledek, který se jen blíží tomu, jak by danou kvalitu vnímal člověk.

Poslechovými testy (všeobecně všemi metodami sloužícími ke stanovení kvality řečového signálu) se zabývá organizace ITU, která stanovuje a zveřejňuje doporučení, jak poslechové testy provádět. Ve dvou následujících kapitolách popíši doporučení ITU-R BS.1116 a ITU-T P.835, které v této práci využiji.

### 3.1.1 ITU-R BS.1116

Standart ITU-R BS.1116 popisuje metodu subjektivního vyhodnocování malých degradací kvality audio signálů vzniklých při použití vysoce kvalitních audio kodérů. Tento standart používá pětistupňovou hodnotící stupnici pro určení míry zkreslení kvality zvukového signálu, tato stupnice je zobrazena v následující tabulce.

Tab. 3.1: Pětistupňová stupnice pro hodnocení kvality zvukového signálu

Význam snížení kvality (Impairment)	Známka (Grade)
Degradace je nepostřehnutelná (Imperceptible)	5
Degradace je slyšitelná, ale neruší (perceptible, but not annoying)	4
Degradace příliš neruší (slightly annoying)	3
Degradace je rušivá (annoying)	2
Degradace je velmi rušivá (very annoying)	1

Tento standard využívá tzv. „Dvojitě slepý A-B-C trojnásobně stimulovaný porovnávací test se skrytou referencí“ („Double blind A-B-C triple-stimulus hidden reference comparison test“). Jsou zde použity tři signály, signál A je signálem referenčním, zatímco signály B a C jsou náhodně vybrány za signálu referenčního a rušeného testovaného signálu. Při poslechovém testu je pak posluchač požádán, aby ohodnotil zhoršení kvality signálu B a signálu C vždy v porovnání se signálem A, a to za využití pětistupňové hodnotící stupnice. Výsledky poslechových testů jsou udávány tzv. subjektivní stupnicí rozdílnosti kvality SDG (Subjective Difference Grade). Hodnota SDG je stanovena odečtením výsledné známky skryté reference od známky příslušného kódovaného signálu:

$$\text{SDG} = \text{ZNÁMKA}_{\text{testovaný signál}} - \text{ZNÁMKA}_{\text{referenční signál}}$$

Pokud vyjde výsledná hodnota SDG rovna nule, pak je testovaný signál bez známky zarušení, pakliže vyjde hodnota SDG rovna -4, tak je testovaný signál silně zkreslen. Testů by se mělo zúčastnit nejméně 20 posluchačů. Výsledky se posléze vyhodnocují statistickou analýzou, vše je popsáno v [4].

### 3.1.2 ITU-T P.835

Doporučení ITU-T P.835 popisuje metodologii pro vyhodnocování řečových systémů, které obsahují algoritmy pro omezení hluku. Toto doporučení však není omezeno jen na tento typ experimentů, rozsah použití je velký. Podrobný popis nároků na účastníky testu, testovacích vzorků a podobně jsou uvedeny v [5], zde je nebudu více rozvádět, popíši jen obecný postup provedení testu.

Průběh poslechového testu podle doporučení ITU-T P.835 je následující: Posluchač si třikrát přehraje testovací signál, při každém poslechu hodnotí jinou oblast zájmu signálu. Při prvním poslechu je pověřen ohodnotit pouze řečový signál podle následující tabulky.

Tab. 3.2: Kvalita řečového signálu

<b>Řečový signál je:</b>	<b>Skóre</b>
Nezdeformovaný (Not distorted)	5
Mírně zdeformovaný (Slightly distorted)	4
Trochu zdeformovaný (Somewhat distorted)	3
Docela zdeformovaný (Fairly distorted)	2
Velmi zdeformovaný (Very distorted)	1

Při druhém poslechu hodnotí rušení pozadí daného vzorku podle následující tabulky.

Tab. 3.3: Rušení pozadí

<b>Pozadí je:</b>	<b>Skóre</b>
Téměř neznatelné (Not noticeable)	5
Mírně znatelné (Slightly noticeable)	4
Znatelné, ale ne rušivé (noticeable but not intrusive)	3
Docela rušivé (Somewhat intrusive)	2
Velmi rušivé (Very intrusive)	1

Při posledním poslechu pak hodnotí celkovou kvalitu signálu.

Tab. 3.4: Celková kvalita signálu

Celková kvalita signálu je:	Skóre
Vynikající (Excellent)	5
Dobrá (Good)	4
Průměrná (Fair)	3
Nízká (Poor)	2
Špatná (Bad)	1

### ***3.2 Metody pro objektivní vnímání kvality zvuku***

Od počátku 90. let 20. století zde byly snahy vyvinout metody pro vyhodnocení vnímané kvality ztrátových širokopásmových technik pro komprese zvuku jako náhrada k drahým poslechovým testům. Mnoho postupů bylo podobných v tom, že používali psychoakustický motivovaný poslechový model, který vyhodnocuje poměr kvality mezi referenčním a testovaným signálem. Simulované poslechové zpracování přetvoří dané signály na odpovídající vnitřní reprezentace, tzn. takové informace, které jsou předpokládány na výstupu poslechového systému (z hlediska nervové aktivity systémů) a slouží vyšším poznávacím úrovním jako vstupy. Tyto vnitřní reprezentace jsou poté zpracovány a porovnány ve smyslu měření matematické vzdálenosti a podobnosti. Tento krok je často označován jako poznávací část modelu. Výsledná měření ideálně korespondují s vnímanými rozdíly signálu vzhledem k referenčnímu signálu, který reprezentuje optimum kvality zvuku.

Predikce vnímaného zkreslení kvality zvuku se ukázala být složitější pro běžné širokopásmové signály s poměrně malým zkreslením než pro nízkopásmovou řeč s výraznějším zkreslením. I když některé postupy objektivního vyhodnocení kvality zvuku fungovaly docela dobře a jeden byl doporučen ITU, tak v roce 1994 žádné

ze sedmi druhů měření kvality nenaplnilo požadavky ITU. V důsledku toho se účastníci dohodli společně vyvinout lepší měřicí metody. Výsledkem byla nová metoda nazývaná percepční hodnocení kvality zvuku (PEAQ = Perceptual Evaluation of Audio Quality), tato metoda je kombinací a rozšířením nejlepších prvků z původních metod a v roce 1998 se stala ITU-R doporučením BS.1378. PEAQ se vyznačuje vysokou mírou optimalizace a adaptace na jediný úkol a představuje realistický a ověřený model sluchového vnímání. Nicméně hlavním účele PEAQ je predikce úrovně kvality nízkobitových kodeků zvukových signálů. Proto byla vytvořena další metoda zvané PEMO-Q, která vychází z metody PEAQ, a kterou popíši v následující kapitole.

### **3.2.1 PEMO-Q**

#### **3.2.1.1 Charakteristika metody**

Metoda PEMO-Q slouží k objektivnímu hodnocení a predikci vnímané kvality zvuku. Metoda je rozšířením měření kvality řeči  $q_c$ , představeným Hansenem a Kollmeierem (PEAQ) a je založena na psychoakusticky ověřeném, kvantitativním modelu efektivního poslechového zpracování. K vyhodnocení zvukové kvality daného zkresleného signálu k odpovídajícímu kvalitnímu referenčnímu signálu se používá poslechový model, který je určen k výpočtu vnitřních reprezentací signálů. Lineární cross-korelace koeficientů vnitřních reprezentací představuje tzv. percepční podobnostní měření PSM (Perceptual similarity measure). PSM ukazuje na dobrou vzájemnou souvislost subjektivního hodnocení kvality v případě, že jsou různé typy signálu uváženy odděleně, zatímco lepší přesnosti predikce kvality nezávislé na signálu je docíleno dalším měřením kvality PMSt reprezentovaným každým 5% ze sekvence okamžité kvality zvuku PMS(t).

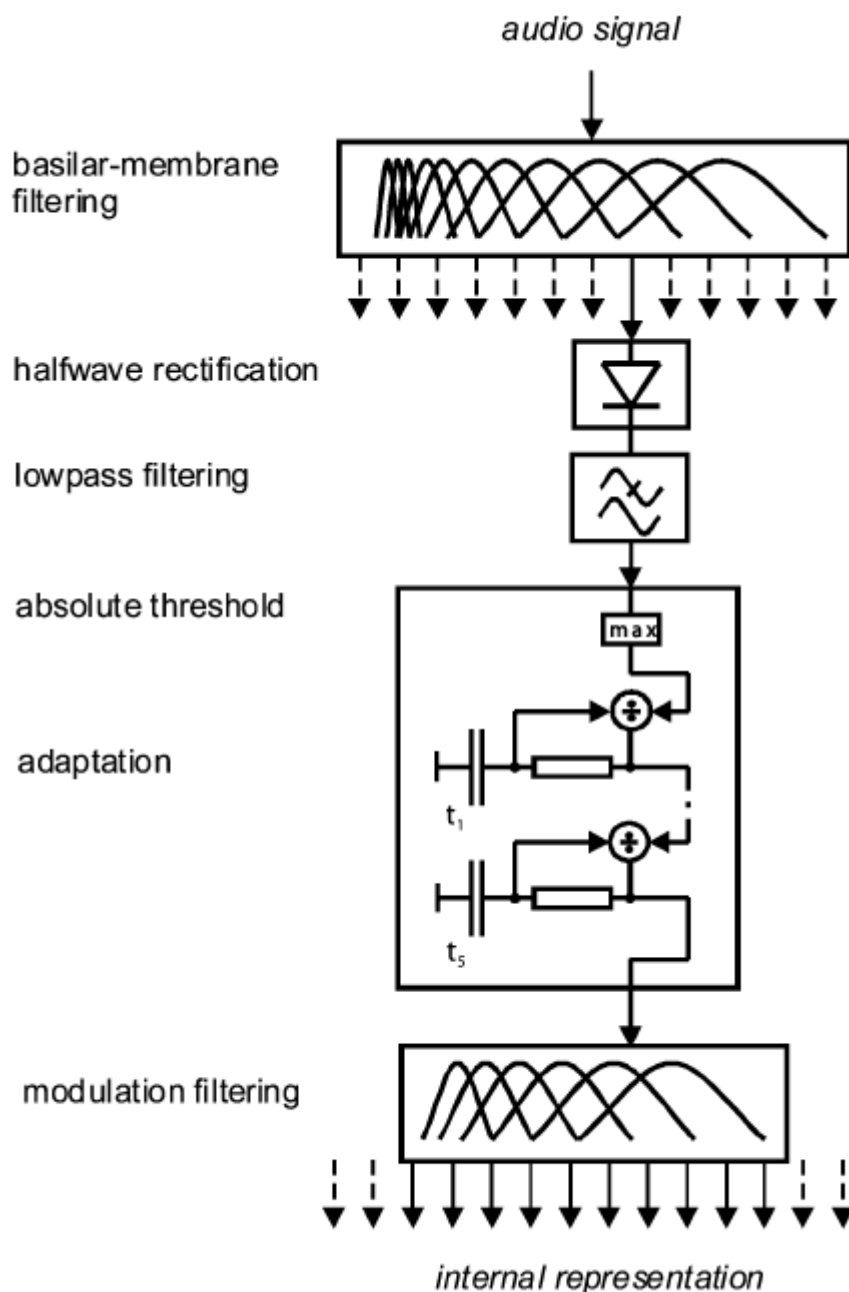
Aby bylo možné testovat a optimalizovat systémy objektivního měření kvality zvuku, byla použita databáze subjektivně hodnocených zvukových signálů s různými typy a stupni zkreslení kvality. Data byla převzata z šesti poslechových testů, které byly provedeny v letech mezi rokem 1990 a 1995 ITU a MPEG (Motion Picture Experts Group) na různých místech. Účelem těchto poslechových testů bylo posoudit kvalitu přenosu různých nízkobitových zvukových kodeků. Na rozdíl od kodeků řeči používaných v telefonní komunikaci, mnoho z testovaných širokopásmových

zvukových kodeků produkovalo podstatně menší poškození. Proto bylo subjektivní hodnocení zpracovaných zvukových signálů prováděno podle ITU-R doporučení BS.1116, které je určeno pro posuzování malých poškození v zvukových systémech. Toto doporučení je popsáno v kapitole 3.1.1.

### **3.2.1.2 Popis metody**

Základní přístup této metody pro objektivní měření kvality zvuku je použít model sluchového zpracování na daný pár referenčního a testovaného signálu a následně vzít korelační koeficienty výstupů modelu jako měřítko pro procentuální podobnost signálů. Předtím než je referenční a testovaný signál zpracovaný poslechovým modelem, musíme tyto signály upravit. Nejprve musíme eliminovat případně časové zpoždění mezi referenčním a testovaným signálem. Tyto odchylky jsou většinou procentuálně irelevantní, ale mohou výrazně ovlivnit objektivní měření kvality. Signály jsou časově zarovnány zpožděním referenčního signálu podle známého nebo odhadnutého zpoždění. Dále musíme provést úroňové zarovnání, které se provádí škálováním testovacího signálu konstantním faktorem. Třetí krok předzpracování spočívá ve vymazání tichých intervalů signálu. Části v referenčním signálu s úrovní výrazně nižší než práh slyšitelnosti jsou vystřižené, stejně tak i odpovídající úseky v testovaném signálu.

Pro simulaci transformace akustického stimulu na nervovou aktivitu lidského ucha se používá kvantitativní model efektivního sluchového zpracování signálu, je použit k předzpracování páru referenčního a testovaného signálu. Tento psychoakusticky motivovaný model transformuje oba příchozí signály na odpovídající vnitřní reprezentace.



Obr. 3.5: Blokové schéma poslechového modelu [1]

Tento obrázek zobrazuje blokové schéma poslechového modelu. Princip jeho fungování nebudu blíže popisovat, je popsán v [1].

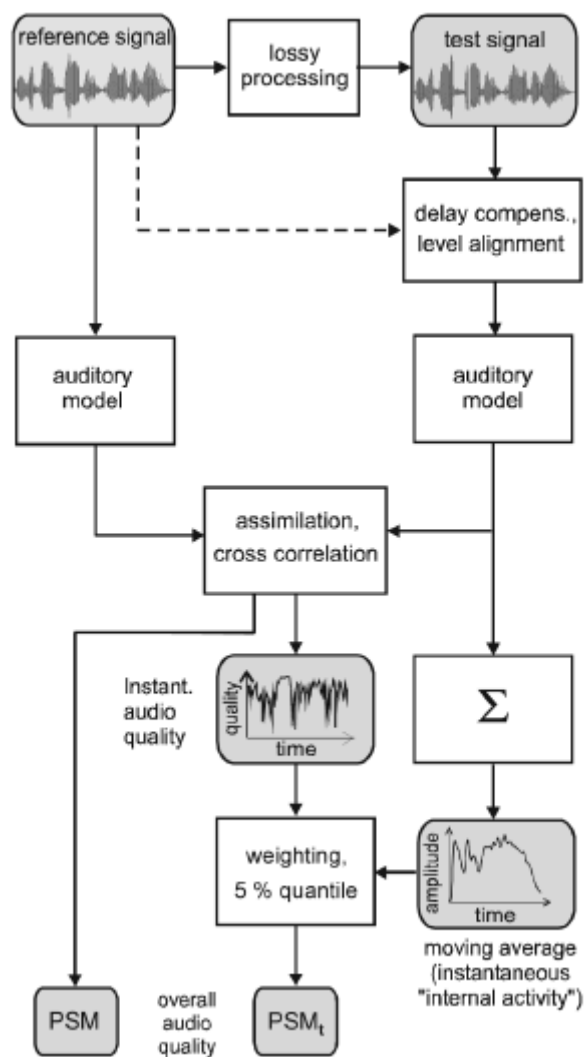
Po zpracování signálu pomocí sluchového modelu následuje výpočet lineární cross-korelace koeficientů interní reprezentace referenčního a testovaného signálu. Nicméně ještě před tímto krokem je vnitřní reprezentace zkresleného testovaného signálu  $Y=(y_{tmf})$  částečně začleněna do referenčního signálu  $X=(x_{tmf})$ . ( $t$  = čas,  $f$  = frekvence,  $m$  = modulace). Prvky z  $Y$ , které mají menší absolutní hodnoty než

odpovídající hodnoty z  $X$ , jsou nahrazeny středními hodnotami obou prvků a tím je rozdíl snížen na polovinu. Závěrečná cross-korelace, která udává vnímané měření kvality (PSM), se provádí samostatně pro každý kanál modulace. (Toto nezávislé zpracování poskytuje vysokou výpočetní účinnost, např. použití různých vzorkovacích frekvencí v rámci jedné vnitřní reprezentace.)

PSM je omezen na interval  $[-1,1]$ , menší hodnoty odpovídají větším odchylkám testovaného signálu od referenčního, což naznačuje zhoršení kvality zvuku testovaného signálu (V praxi nejsou negativní hodnoty PSM pozorovány.) Navíc se počítá okamžitá subjektivní kvalita zvuku  $PMS_{(t)}$ . Následně se bere každé páté procento z  $PMS_{(t)}$  a slouží k výpočtu druhého měření celkové kvality zvuku, nazývané  $PSM_t$ . Tato měření ukazují vztah mezi okamžitým a celkovým vnímáním kvality zvuku. Hlavním rozdílem mezi oběma měřeními je takový, že  $PSM_t$  nezáleží na typu vstupního signálu, v kontrastu s PSM. Aby bylo možné předpovědět subjektivní hodnocení podle stupnice SDG (Subjective Difference Grade), je  $PSM_t$  mapováno na stupnici ODG (Objective Difference Grade) pomocí regresní funkce.

Pro lepší představu o činnosti metody PEMO-Q si můžete prohlédnout její blokové schéma (Obr. 3.6: Blokové schéma metody PEMO-Q).





Obr. 3.6: Blokové schéma metody PEMO-Q

Metoda PEMO-Q představuje mnohem přímočařejší přístup s mnohem jednodušší poznávací částí modelu než u ITU-R doporučení BS. 1387 (PEAQ). Kvůli nižšímu stupni specializace má metoda PEMO-Q vyšší schopnost zobecňovat a tím se stává aplikovatelnou na neznámá narušení a zvukový materiál.

## 4 Hlasitost signálu

Jedním z důležitých aspektů, který ovlivňuje kvalitu zvukového signálu, je jeho hlasitost. Hlasitost je subjektivní veličina, která je úměrná energii signálu. energii signálu, respektive jeho výkon, definujeme následujícím vztahem:

$$E = \frac{1}{N} \sum_{i=0}^N x[i]^2,$$

V praxi se pak více používá logaritmická stupnice s jednotkou decibel [dB], čehož dosáhneme zlogaritmováním předchozího vztahu:

$$E = 10 \cdot \log_{10} \frac{1}{N} \sum_{i=1}^N x[i]^2$$

Lidské ucho však vnímá hlasitost zvuku složitěji. Jedním z faktorů, který ovlivňuje hlasitost zvuku, je jeho frekvence. Mezi aparáty na zpracování zvuku, které berou tuto frekvenční závislost v potaz, patří váhový filtr A a křivky konstantní hlasitosti.

### 4.1 Váhový filtr A

Lidské ucho je různě citlivé na různé frekvence zvuku. Váhový filtr A funguje tak, že upravuje citlivost zvukoměrů, aby byla podobná vnímání lidského ucha. Zvukoměry se zabudovanými váhovými filtry A se používají k měření v pracovním prostředí a ve venkovních prostorech. Existují i jiné typy váhových filtrů (B, C, D) například filtr C se používá při měření zvukových vlastností chráničů sluchu nebo při vědeckých měřeních. Váhový filtr A je vlastně konkrétním zjednodušením křivek konstantní hlasitosti popsaných níže. [3]

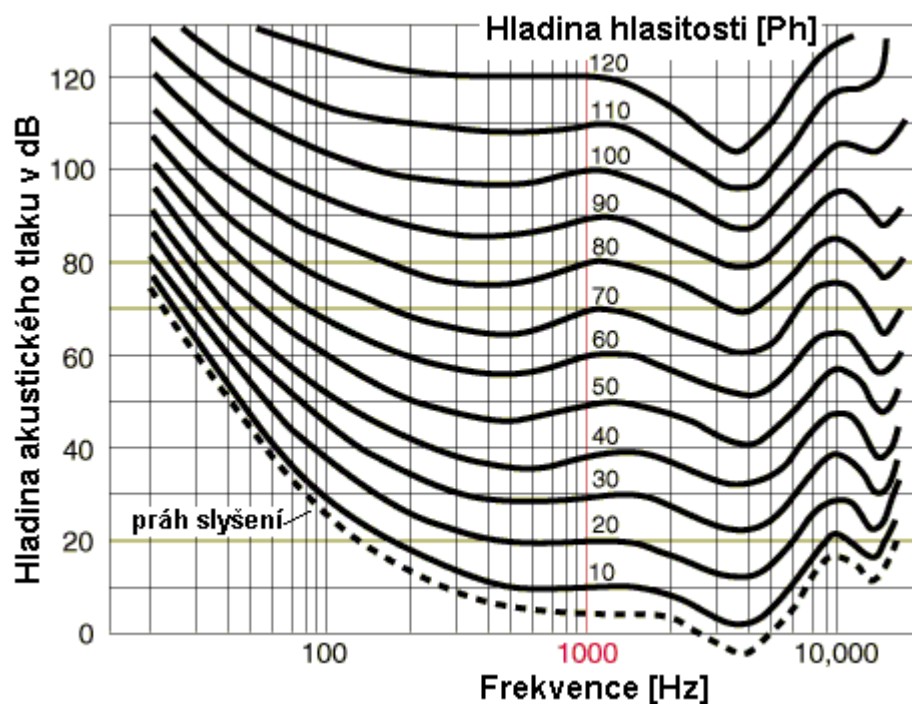
V následující tabulce si můžete prohlédnout korekce po použití váhového filtru A.

Tab. 4.1: Tabulka korekcí váhového filtru A [11]

Frekv. [Hz]	korekce [dB]	Frekv. [Hz]	korekce [dB]	Frekv. [Hz]	korekce [dB]
20	-50,5	200	-10,9	2 k	+1,2
25	-44,7	250	-8,6	2,5 k	+1,3
31,5	-39,4	315	-6,6	3,15 k	+1,2
40	-34,6	400	-4,8	4 k	+1,0
50	-30,2	500	-3,2	5 k	+0,5
63	-26,2	630	-1,9	6,3 k	-0,1
80	-22,5	800	-0,8	8 k	-1,1
100	-19,1	1 k	-0,0	10 k	-2,5
125	-16,1	1,25 k	+0,6	12,5 k	-4,3
160	-13,4	1,6 k	+1,0	16 k	-6,6
				20 k	-9,3

## 4.2 Křivky konstantní hlasitosti (Equal-loudness)

Dalším aparátem, který upravuje hlasitost zvuku podle vnímání lidského ucha, jsou tzv. křivky konstantní hlasitosti (Obr: Equal-loudness křivky). Míra citlivosti sluchu způsobená zvukem není přímo úměrná fyzikální energii. Fechner-Weberův zákon nám říká, že hlasitost roste s logaritmem intenzity zvuku, tato závislost však není lineární, ale je frekvenčně závislá. Jednotkou hlasitosti je fón a je definován tak, že hladina hlasitosti 1 fón je při kmitočtu 1kHz stejně velká jako jednotka hladiny zvuku 1dB.



Obr. 4.2: Equal-loudness křivky [10]

Tyto křivky jsou lepší aproximací k citlivosti lidského sluchu než A-křivka a byly určeny experimentálně. Každá křivka odpovídá jedné hladině hlasitosti. Prohlédneme-li si průběh křivek konstantní hlasitosti, zjistíme, že hodnoty intenzity v dB a hlasitosti ve fónech si odpovídají jen pro frekvenci 1 kHz. U ostatních frekvencí nám křivky udávají to, jaký akustický tlak musí zvuk mít, aby u člověka vyvolal stejně hlasitý vjem jako při frekvenci 1kHz. [3]

## 5 Databáze řečových signálů

Tato kapitola obsahuje popis vytvoření databáze testovacích signálů. Cílem je vytvořit řečové vzorky s různými typy šumu v pozadí signálů a s různými zkresleními. Tyto vzorky budou posléze sloužit jako testovací vzorky poslechových testů, metody PEMO-Q a k vyhodnocení hlasitosti.

Vzorky řečového signálu musejí být jednoduché, snadno pochopitelné, zároveň však musí obsahovat různé rozmanitosti řeči. Já jsem postupoval tak, že jsem vybral vhodné věty z článků internetových magazínů (jejich výpis naleznete na přiloženém cd). Tyto věty pak byly zaznamenány standardním externím mikrofonom s důrazem na plynulou výslovnost a absenci řečových nedostatků. Mluvčí byl čtyři, dva muži (věk 25 a 49) a dvě ženy (věk 22 a 45).

K zašumění těchto čistých řečových signálů byly použity signály šumů z [15]. Mezi stacionární šумы patří bílý a technický šum a zvuk vysavače, jako nestacionární šum pak byl použit hluk davu lidí, hluk z křižovatky a zvuk jedoucího auta s měnícími se otáčkami motoru. Stacionární signály jsou takové, u kterých se statistické vlastnosti (jako je třeba spektrum signálu, nebo jeho hlasitost) nemění v čase. U nestacionárních signálů ke změnám statistických vlastností v čase dochází. Zašumění, neboli sečtení čistého řečového signálu a šumu jsem provedl v programu Matlab podle rovnice:

$$y[n] = s[n] + k \cdot n[n],$$

kde  $y[n]$  je výsledný zašuměný signál,  $s[n]$  je čistý řečový signál,  $n[n]$  je šum a  $k$  je koeficient upravující poměr signálu a šumu.

Hodnota koeficientu  $k$  se odvíjí od hodnoty SNR (Signal-to-Noise Ratio), která nám vyjadřuje odstup signálu od šumu. Výpočet SNR se provádí podle následujícího vzorce:

$$SNR = 10 \cdot \log \frac{\sigma_s^2}{\sigma_n^2},$$

kde  $\sigma_s^2$  je výkon užitečného signálu a  $\sigma_n^2$  je výkon šumu. Chceme-li ale míchat šum a užitečný signál v nějakém poměru, musíme rovnici upravit přidáním konstanty  $k$ , rovnice pak vypadá následovně:

$$SNR = 10 \cdot \log \frac{\sigma_s^2}{k^2 \cdot \sigma_n^2}.$$

Při přípravě testovacích signálů však vycházíme z předpokladu, že SNR známe, respektive si ho zvolíme na námi požadovanou úroveň, a koeficient  $k$  si poté dopočítáme podle vzorce, který odvodíme z předchozího vztahu:

$$k = \sqrt{\frac{\sigma_s^2}{\sigma_n^2}} \cdot 10^{-\frac{SNR}{10}}.$$

Konkrétně jsem si tedy vytvořil 30 řečových signálů namluvených dvěma muži a dvěma ženami. Ke každému signálu jsem posléze přičetl šum a to tak, že první skupina 30 vzniklých testovacích vzorků obsahovala stacionární šumy (na každý jednotlivý šum tedy připadlo 10 vzorků), druhá skupina obsahovala stejné řečové vzorky ale s přičtenými nestacionárními šumy (opět na každý jednotlivý šum tedy připadlo 10 vzorků). Poměr odstupe signálu od šumu SNR u vzniklých vzorků je roven hodnotám -2dB, 4dB a 10dB, což zaručuje zahrnutí běžných úrovní šumů do testů.

Dále jsem řečové signály upravil pomocí dvou filtrů, a to jednu třetinu pomocí pásmové propusti, která propouštěla frekvence v rozmezí 300Hz – 3400Hz. Tento filtr simuluje telefonní hovor. Druhou třetinu řečových signálů jsem zfiltroval pásmovou zádrží, která nepropouštěla frekvence v rozmezí 500Hz – 1000Hz, což je oblast, která ovlivňuje kvalitu řeči. Poslední třetinu řečových signálů jsem ponechal beze změn. Tabulku všech testovacích signálů naleznete na příloženém cd (*databaze/prehled.xml*).

## 6 Vyhodnocení poslechových testů

Hlavním cílem těchto poslechových testů bylo vyhodnotit databázi testovacích signálů. Tedy zhodnotit vliv míry zastoupení aditivního šumu na kvalitu řečového signálu, porovnat rušení stacionárních a nestacionárních šumů, zjistit vliv průchodu signálu filtrem na jeho kvalitu. Dosažené výsledky pak budou sloužit jako absolutní měřítko pro vyhodnocení funkčnosti metody PEMO-Q.

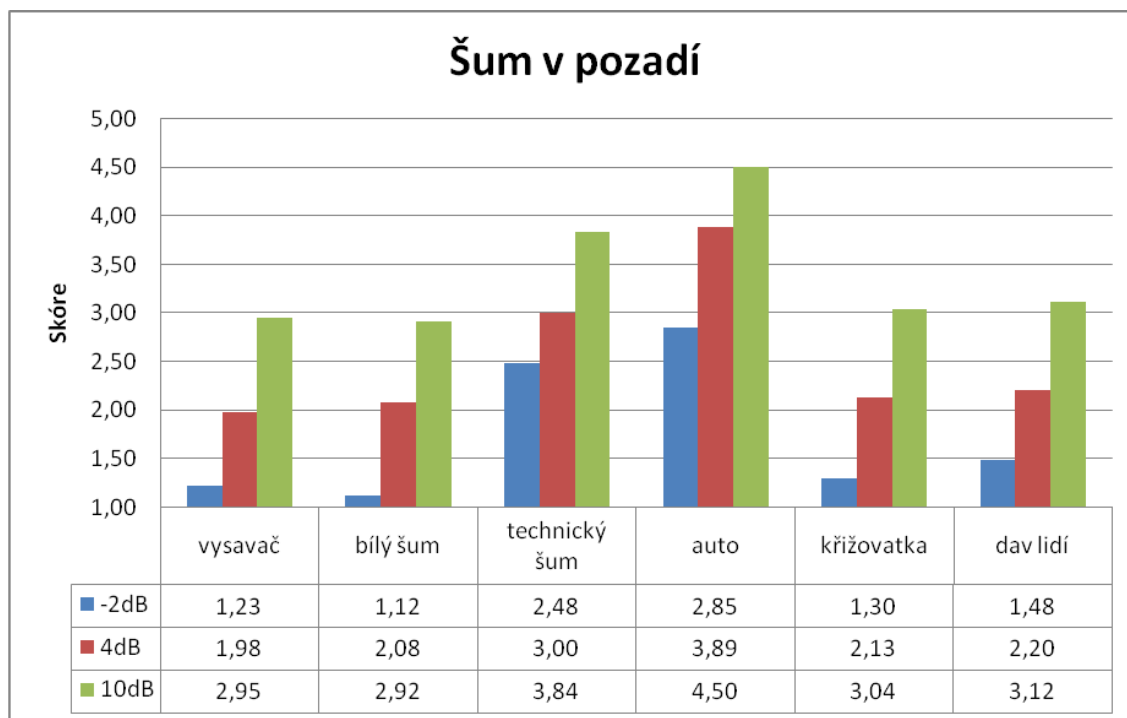
Poslechové testy byly provedeny podle doporučení ITU-T P.835, které popisuje metodologii pro vyhodnocování řečových systémů, které obsahují algoritmy pro omezení hluku. Popis této metodologie naleznete v kapitole 3.1.2. Poslechových testů se zúčastnilo 20 posluchačů, kteří hodnotili databázi řečových vzorků, obsahující 60 testovacích vzorků s různými druhy šumu v pozadí a s různými odstupy signálu od šumu SNR. Popis vytvoření testovací databáze řečových vzorků naleznete v kapitole 5.

Instrukce, které obdržel každý účastník testu, naleznete v Příloze A. Posluchači hodnotili tři oblasti testovacího signálu, nejprve kvalitu vlastního řečového signálu, dále rušení aditivního šumu v pozadí signálu a nakonec celkovou kvalitu signálu. V následujících kapitolách popíši výsledky jednotlivých částí poslechových testů, začnu s popisem rušení šumu v pozadí signálu, která výrazně ovlivňuje zbývající dvě měření.

### 6.1 Šum v pozadí signálu

Hodnocení se provádělo podle tabulky (Tab. 3.3: Rušení pozadí) zobrazené v kapitole 3.1.2, tedy stupnice se pohybuje v rozmezí hodnot skóre 5 (pozadí je téměř neznatelné) až po hodnotu skóre 1 (pozadí je velmi rušivé). V této tabulce jsem provedl jedinou změnu oproti doporučení ITU-T P.835, kde hodnota skóre 5 je popsána jako neznatelné pozadí. Jelikož v testovací databázi, na které byly prováděny poslechové testy, nejsou signály, které by neobsahovaly žádný šum, upravil jsem popis tohoto skóre jako téměř neznatelné pozadí, aby v poslechových testech byla využita celá stupnice.

Výsledky rušení šumů v pozadí signálů jsou zobrazeny v následujícím grafu (Graf 6.1: Šum v pozadí). Každá skupina tří sloupců zastupuje jeden druh šumu, jednotlivé sloupce pak představují v jakém poměru SNR je daný šum v signálu zastoupen (tedy -2dB, 4dB a 10dB). První tři šumy se řadí mezi stacionární, čtvrtý až šestý pak mezi nestacionární.



Graf 6.1: Šum v pozadí

Na první pohled je z grafu patrná závislost rušení šumu na jeho poměrném zastoupení v signálu, čili pro hodnotu SNR rovnající se -2dB (tzn. šum je o 2dB hlasitější než řeč) je rušení výraznější než pro hodnotu SNR 4dB, a to je opět více rušivé než u SNR odpovídajícímu 10dB, kdy už je řeč o 10 dB hlasitější než šum. Odstupy mezi hodnotami skóre pro -2dB a 4dB a pro 4dB a 10dB se od sebe liší velmi málo, z čehož lze usuzovat, že závislost rušení šumu na jeho poměrném zastoupení v signálu je téměř lineární.

Porovnáme-li rušení stacionárních a nestacionárních šumů, je patrné, že nestacionární šumy jsou o něco méně rušivé při stejných poměrech zastoupení v celkovém signálu, což dokazuje i následující tabulka.



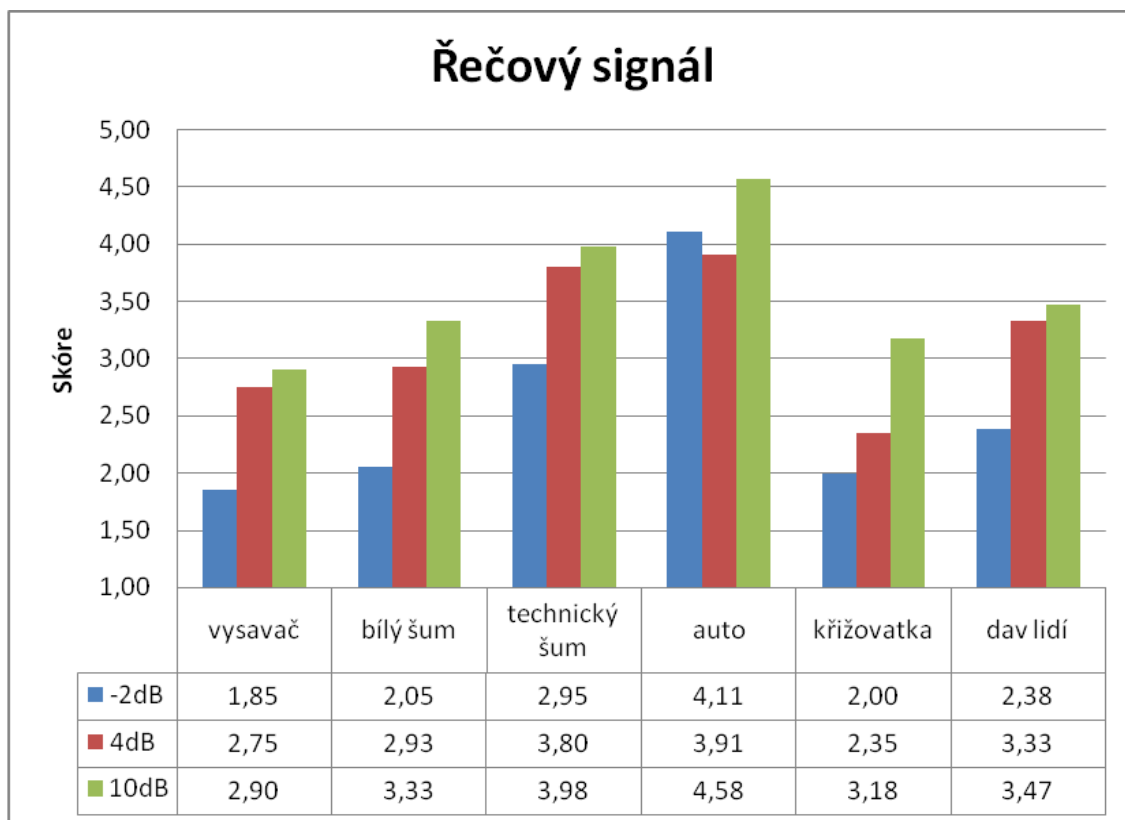
Tab. 6.2: Porovnání stacionárních a nestacionárních šumů

	-2dB	4dB	10dB
Stacionární šumy	1,61	2,35	3,23
Nestacionární šumy	1,88	2,74	3,55

U stacionárních šumů výrazně vybočuje rušení technického šumu, u nestacionárních se značně liší rušení zvuku auta, oba šumy jsou méně rušivé než ostatní.

## 6.2 Kvalita řečového signálu

Hodnocení se provádělo podle tabulky (Tab. 3.2: Kvalita řečového signálu) zobrazené v kapitole 3.1.2, tedy stupnice se opět pohybuje v rozmezí hodnot skóre 5 (řečový signál je nezdeformovaný) až po hodnotu skóre 1 (řečový signál je velmi zdeformovaný). Může se zdát, že na kvalitu řečového signálu, čili na jeho zdeformovanost, bude mít vliv pouze průchod signálu nějakým filtrem. Přidáním šumu do pozadí signálu by už pak kvalitu vlastního řečového signálu ovlivnit nemělo, což ovšem poslechové testy nepotvrdili. Jejich výsledky si můžete prohlédnout v grafu (Graf 6.3: Řečový signál).

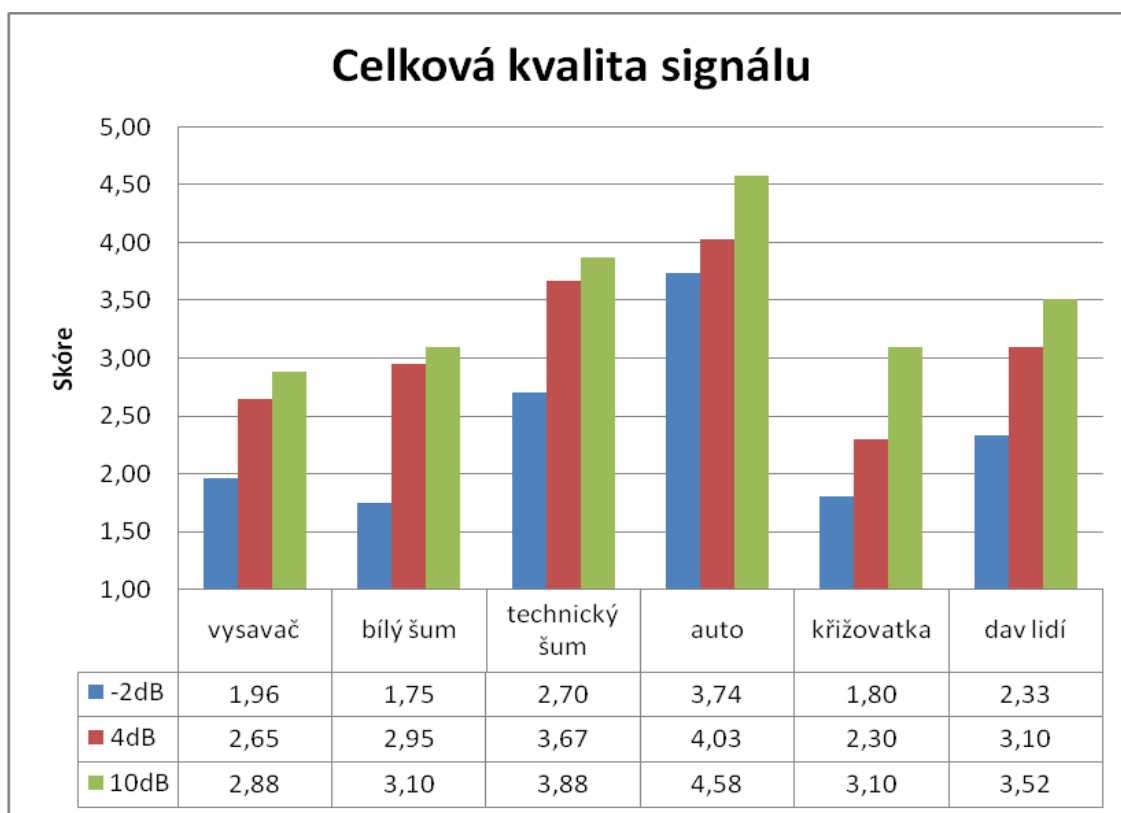


Graf 6.3: Řečový signál

Jak je vidět, kvalita řečového signálu je výrazně ovlivněna mírou zastoupení aditivního šumu. Jediným případem, který tento fakt nesplňuje, jsou signály rušené zvukem auta (signály se šumem zastoupeném v signálu v poměru SNR 4dB jsou méně zdeformované než u SNR -2dB). Tento šum však z testu vyšel jako nejméně rušivý, což zapříčinilo to, že nejméně ovlivňuje kvalitu řečového signálu. Proto je i na tomto případě dobře patrný vliv filtru na kvalitu řečového signálu. Signály rušené hlukem auta v poměru SNR -2dB byly v testu čtyři, z toho tři neprošli žádným filtrem a jeden byl filtrován filtrem simulující hovorový kanál. Tento filtrovaný signál byl v poslechových testech hodnocen průměrným skóre 3.90, zatímco nefiltrované signály měly průměrné skóre 4.15 respektive 4.20, čili vykazovali lepší kvalitu. Signály rušené hlukem auta v poměru SNR 4dB byly v testu také čtyři, tři z nich prošly filtrem pásmová zádrž (hodnoty skóre 3.90, 3.65 a 3.90) a jeden pásmovou propustí simulující hovorový kanál (hodnota skóre 4.20). Proto je u těchto signálů kvalita řečového signálu horší i přesto, že jsou rušeny slabším šumem. Na to jak, daný filtr ovlivní kvalitu řeči, má také výrazný vliv samotný řečník. Každý člověk má hlas na jiné frekvenční úrovni, proto je i ovlivnění filtrem subjektivní.

### 6.3 Celková kvalita signálu

Posledním oblastí zájmu hodnocení byla celková kvalita signálu, jako souhrn kvality řečového signálu a rušení pozadí, čili celkový vjem ze signálu a jeho srozumitelnosti. Hodnocení probíhalo podle tabulky (Tab. 3.4: Celková kvalita signálu) zobrazené v kapitole 3.1.2. Tedy podle stupnice v rozmezí od skóre 5 (celková kvalita signálu je vynikající) až po skóre 1 (celková kvalita signálu je špatná). Výsledky jsou zobrazené v grafu (Graf 6.4: Celková kvalita signálu).



Graf 6.4: Celková kvalita signálu

Opět je zde vidět výrazná závislost na hladině aditivního šumu a z toho vyplývající lepší kvalita signálu rušených šumy auto a technickým šumem, všeobecně pak lepší kvalita signálu s nestacionárními šumy.

Tab. 6.5: Porovnání stacionárních a nestacionárních šumů

	-2dB	4dB	10dB
Stacionární šumy	2,14	3,09	3,29
Nestacionární šumy	2,62	3,14	3,73

U stacionárních šumů je patrné, že mezi hladinami šumu SNR 4dB a 10dB není v celkové kvalitě signálu tak velký rozdíl oproti hladině -2dB, kdy je celková kvalita výrazně snížena.

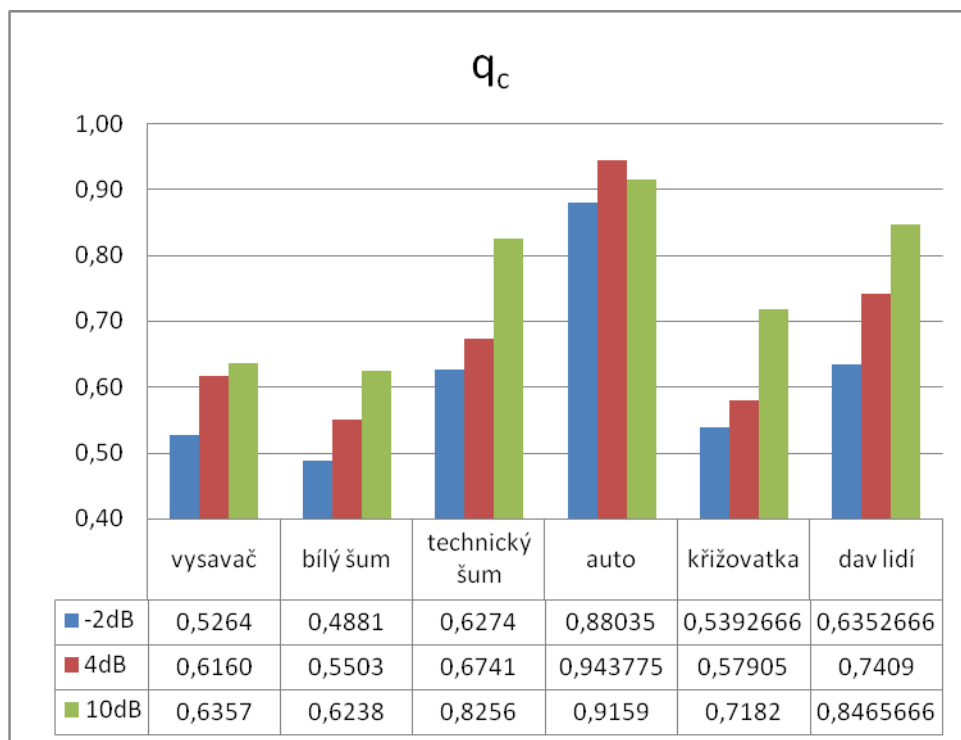
## 7 Vyhodnocení objektivní metody PEMO-Q

Metoda PEMO-Q slouží k objektivnímu hodnocení a predikci vnímané kvality zvuku. Tuto metodu jsem testoval pomocí programu PEMO-Q v.1.3.1, což je aplikace ovládaná z příkazového řádku operačního systému Windows. Aplikace obsahuje dva podprogramy, prvním z nich je *speechqual.exe*, který počítá objektivní kvalitu řeči  $q_c$ , druhý podprogram *audioqual.exe* je pak už vlastní implementací metody PEMO-Q a vrací hodnoty PSM a  $PSM_t$ . Popis a výsledky naleznete v následujících kapitolách.

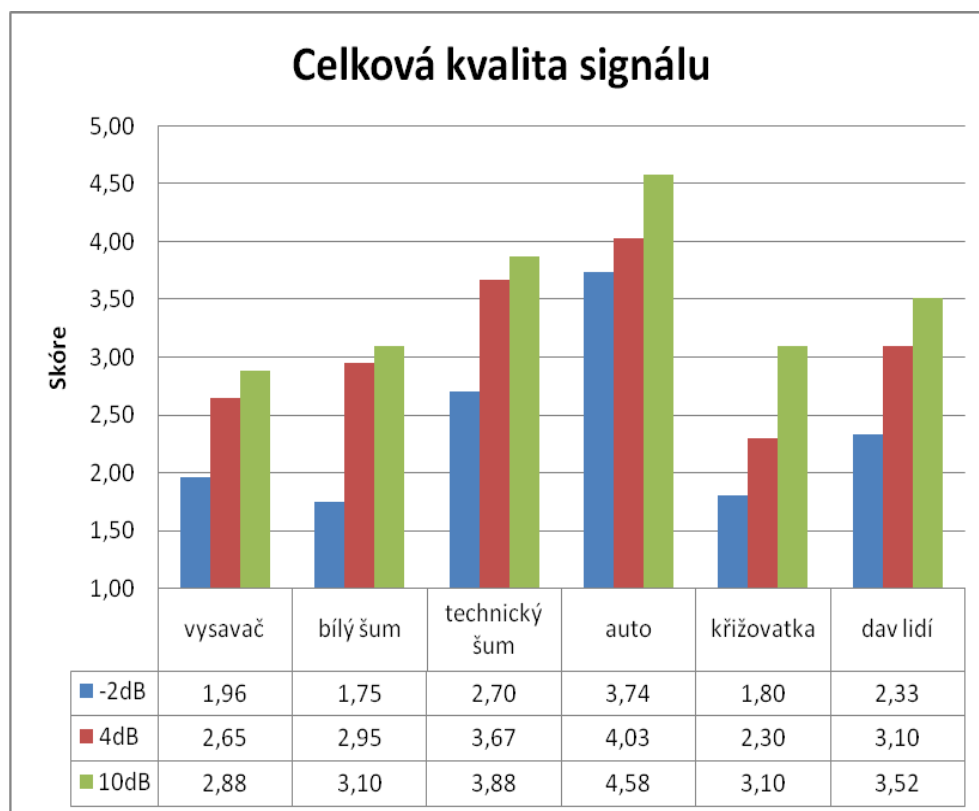
### 7.1 Objektivní kvalita řeči $q_c$

Program *speechqual.exe* je implementací metody představené pány Hansen a Kollmeier (PEAQ), ze které metoda PEMO-Q vychází. Metoda Hansena a Kollmeiera byla původně navržena pro predikci kvality telefonních hovorů s relativně malým zkreslením způsobeným nízkobitovými řečovými kodeky. Metoda předvídá vnímanou kvalitu daného testovaného řečového signálu v porovnání s odpovídajícím referenčním signálem, za použití modelu poslechového vnímání. Výstupem této metody je objektivní kvalita řeči  $q_c$ , což reprezentuje vnímanou podobnost testovaného a referenčního signálu. Pokud má referenční signál vysokou kvalitu, pak můžeme  $q_c$  pokládat za objektivní měření kvality řeči testovaného signálu.

Výsledky dosažené pomocí výše popsané metody jsou zobrazeny v grafu (Graf 7.1: Objektivní kvalita řeči  $q_c$ ). Kvůli lepšímu porovnání bude následovat i graf zobrazující celkovou kvalitu signálu z poslechových testů (Graf 7.2: Celková kvalita signálu). Čím více se hodnota  $q_c$  blíží hodnotě jedna, tím je testovaný signál podobnější referenčnímu, tím pádem je kvalitnější.



Graf 7.1: Objektivní kvalita řeči  $q_c$



Graf 7.2: Celková kvalita signálu

Porovnáme-li dosažené výsledky s výsledky poslechových testů (respektive s částí Celková kvalita signálu) jsou zde patrné výrazné společné znaky. Prvním z nich je závislost na míře aditivního šumu obsaženém v signálu. Tato závislost je porušena pouze v případě signálu s šumem auto na úrovni SNR 10dB, zde však musím zdůraznit, že tyto signály byly pouze dva, první byl ohodnocen  $q_c=0.9798$  a druhý však pouze  $q_c=0.852$ . Toto snížení kvality přisuzuji tomu, že signál byl namluven mluvčím označeným Muž2, jehož řečové signály dosahovaly v tomto testu celkově horších výsledků než u ostatních mluvčích. Dalším společným znakem je zvýšená kvalita signálů s nestacionárními šumy, což je patrné v následující tabulce.

Tab. 7.3: Porovnání stacionárních a nestacionárních šumů

	-2dB	4dB	10dB
Stacionární šumy	0,55	0,61	0,70
Nestacionární šumy	0,68	0,75	0,83

Celkově se dá říci, že metoda vytvořená pány Hansen a Kollmeier implementovaná do programu *speechqual.exe* dává takové výsledky, které se v důležitých aspektech shodují s výsledky, které vyplynuly ze subjektivních poslechových testů. Přesnějších výsledků by ale měla dosahovat metoda PEMO-Q.

## 7.2 Měření vnímané podobnosti PSM

Program *audioqual.exe* je implementací metody pro objektivní percepční hodnocení kvality zvuku PEMO-Q představené pány Huber a Kollmeier v roce 2006. Důkladný popis metody naleznete v kapitole 3.2.1. Metoda PEMO-Q vychází z metody PEAQ, metoda opět porovnává testovaný a referenční signál a jejím prvním výstupem je hodnota měření vnímané podobnosti PSM (Perceived Similarity Measure), která svým způsobem odpovídá objektivní kvalitě řeči  $q_c$ . Hodnotu PSM je opět za podmínky vysoké kvality referenčního signálu možno brát jako objektivní měření kvality řeči testovaného signálu. Metoda navíc počítá hodnotu okamžité kvality zvuku  $PSM(t)$ , která slouží k výpočtu celkového měření kvality  $PSM_t$ , což je druhý výstup metody PEMO-Q. Obě měření se pohybují v rozsahu  $[-1, 1]$  (v praxi se však záporné

hodnoty u měření PSM nevyskytují), přičemž čím víc se měření blíží hodnotě 1, tím je testovaný signál kvalitnější.

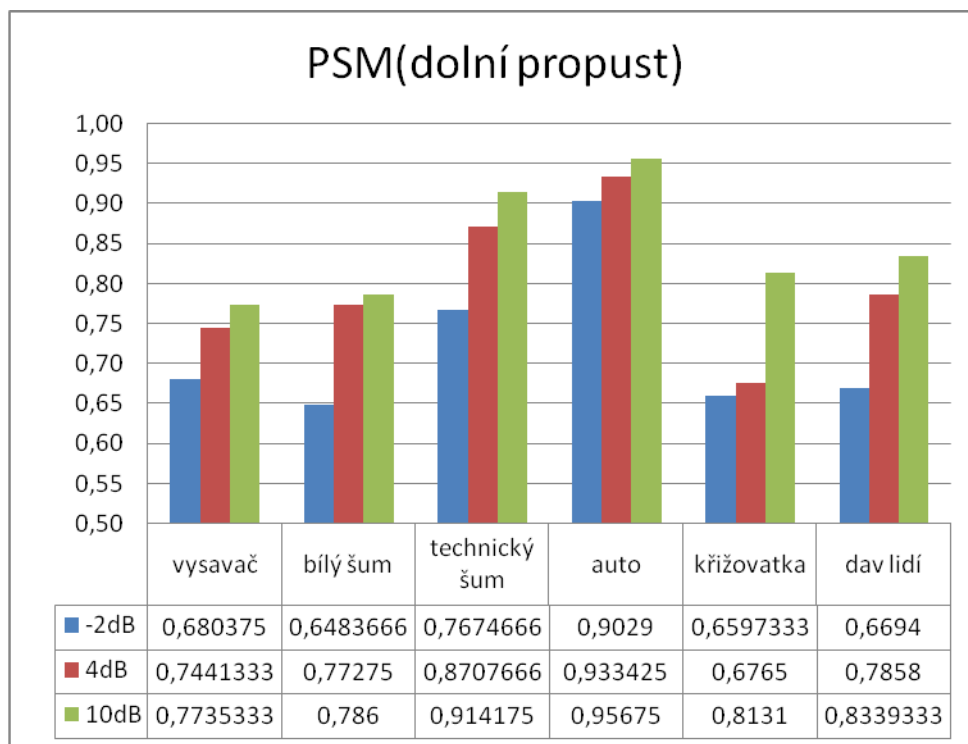
Podle podkladů dodaných k metodě PEMO-Q [16] by mělo  $PSM_t$  dosahovat lepších výsledků při vyhodnocování zvukových kodeků (čili u malých zkreslení zvukového signálu), zatímco v ostatních oblastech vyhodnocování zvukových signálů, jako je třeba testovací databáze použitá v této práci, je použitelnější prostší měření PSM. Takže prvním předpokladem je, že měření PSM by nám mělo dávat na zde použité testovací databázi více odpovídající výsledky než měření  $PSM_t$ .

Druhý předpoklad se týká modulačního procesu v percepčním modelu, kde máme na výběr mezi dvěma možnostmi, buď rychlejším filtrem dolní propust, nebo pomalejší filtr banky, která by měla dosahovat lepších výsledků. Nicméně v [16] je uvedeno, že při použití filtr banky, je lepších výsledků dosaženo jen v případě malých zkreslení kvality testovaných zvukových signálů, což není náš případ.

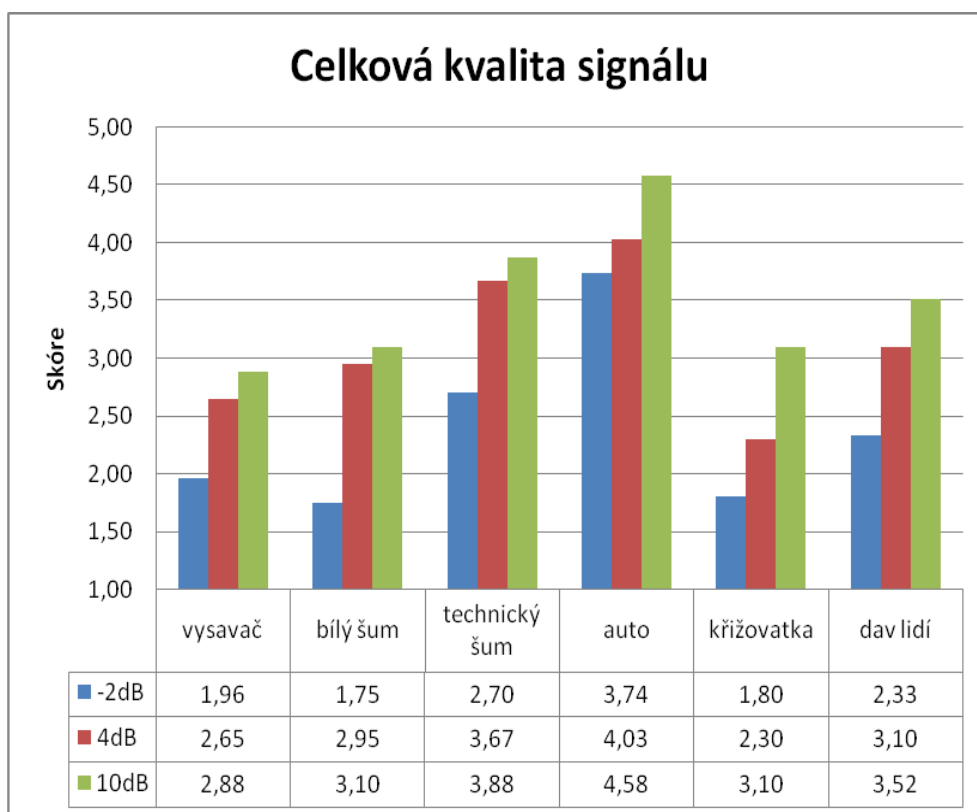
Celkově jsem tedy metodu PEMO-Q otestoval ve 4 variantách, tedy měření PSM s filtrem dolní propust a filtr bankou a stejně tak měření  $PSM_t$ . Abych mohl vyhodnotit, která z těchto variant dosahuje lepších výsledků, budu ji porovnávat s výsledky poslechových testů provedených na stejné databázi řečových signálů. Poslechové testy tedy považuji za absolutní měřítko pro posouzení kvality zvuku.

Co se týče výběru modulačního procesu, daleko více odpovídajících výsledků bylo dosaženo pomocí jednodušší dolní propusti. K daleko lepším výsledkům musíme ještě přičíst kratší čas výpočtu, který se při použití dolní propusti pohyboval v řádech jednotek sekund, oproti tomu při použití filtr banky vzrostl výpočetní čas až na desítky sekund. Výsledky měření PSM (Graf 7.3: PSM) a  $PSM_t$  (Graf 7.5:  $PSM_t$ ) při použití dolní propusti si můžete prohlédnout v grafech. (Výsledky měření PSM a  $PSM_t$  při použití filtr banky si můžete v souboru *PEMO-Q/vysledky.xml* na přiloženém CD.) Kvůli lepšímu porovnání bude následovat i graf zobrazující celkovou kvalitu signálu z poslechových testů (Graf 7.5: Celková kvalita signálu).





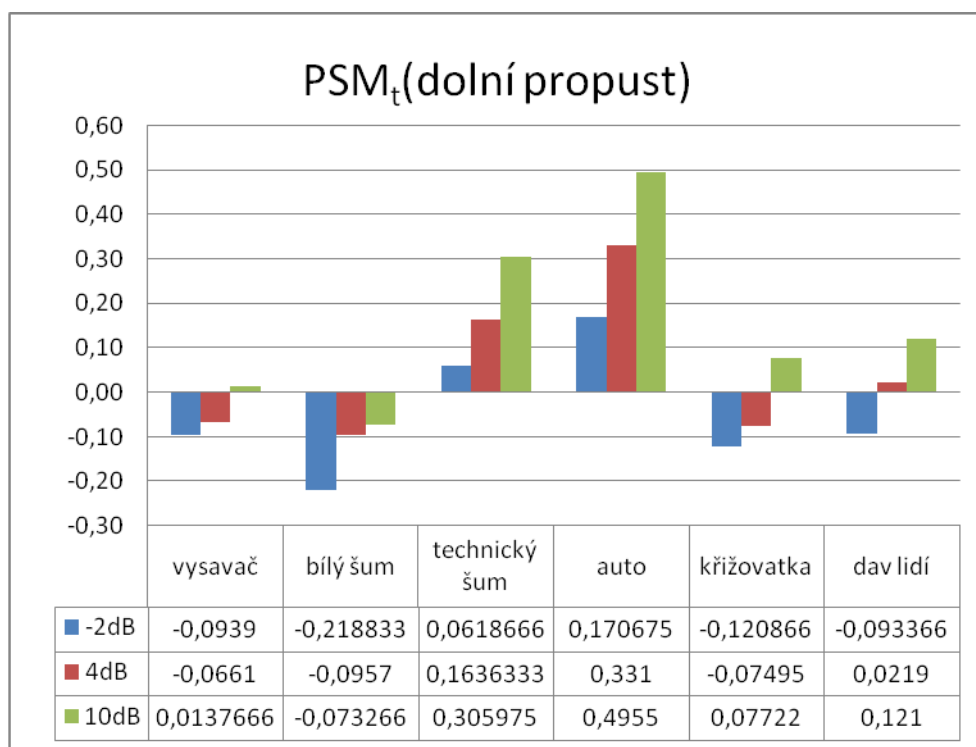
Graf 7.4: PSM



Graf 7.5: Celková kvalita signálu

Tab. 7.6: Porovnání stacionárních a nestacionárních šumů

	-2dB	4dB	10dB
Stacionární šumy	0,70	0,80	0,82
Nestacionární šumy	0,74	0,80	0,87



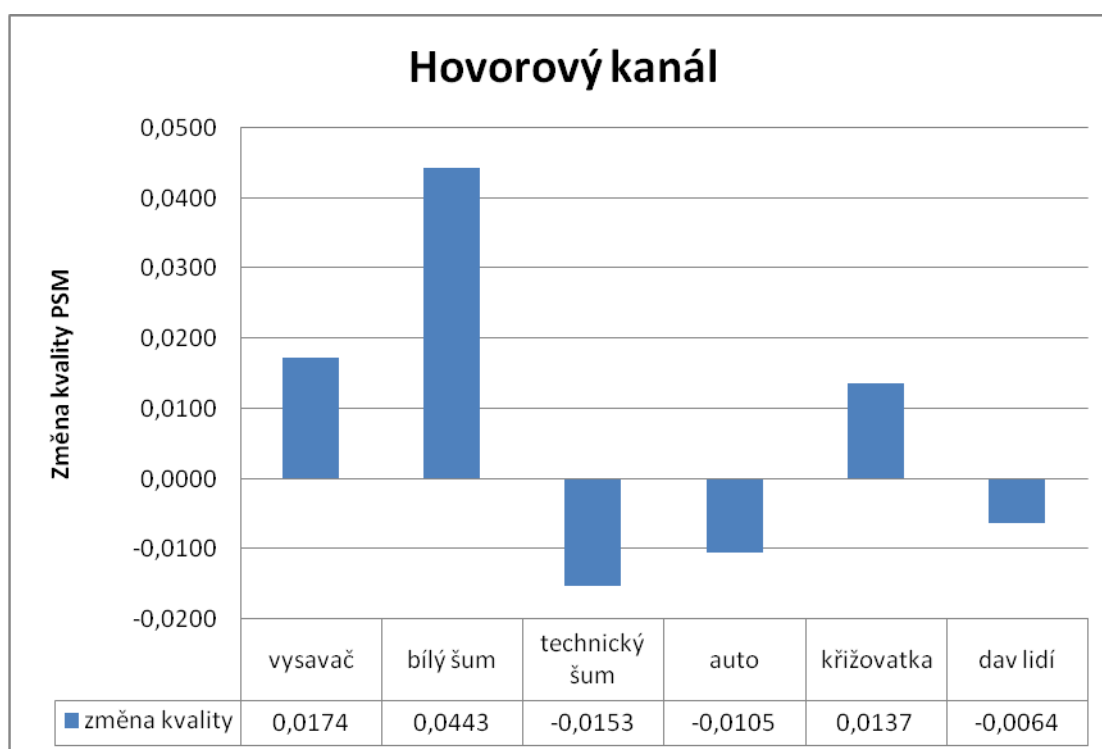
Graf 7.7: PSM<sub>t</sub>

U obou měření si výsledky zachovaly stejné hlavní rysy jako u výsledků z poslechových testů, jako je výrazná závislost kvality signálů na míře zastoupení aditivního šumu, lepší kvalita signálů s šumem auto a technickým šumem, o něco lepší kvalita signálů s nestacionárními šumy (Tab. 7.4: Porovnání stacionárních a nestacionárních šumů).

Na první pohled je patrné výrazné snížení kvality u měření PSM<sub>t</sub>, což je způsobeno tím, že toto měření se daleko více hodí na vyhodnocování zvukových kodeků, čili daleko menších zkreslení kvality. Proto nejpřesnějším měřením pro vyhodnocení řečových signálů se silným rušením šumem je měření PSM při použití filtru dolní propust.

Měření PSM jsem využil k vyhodnocení toho, jak se změní kvalita signálu po aplikaci filtru simulující hovorový kanál a pásmové zádrže. Čili porovnal jsem kvalitu řečových signálů s aditivním šumem s odpovídajícím signálem, který prošel jedním ze zmíněných filtrů. Filtr pásmová zádrž v rozmezí 500 – 1000Hz u všech signálů snížil kvalitu a to o průměrnou hodnotu PSM 0,037, což je způsobeno tím, že filtr ovlivňoval kvalitu řečového signálu.

Pásmová propust simulující hovorový kanál ovlivňovala kvalitu v závislosti na aditivním šumu, což je patrné z následujícího grafu (Graf 7.6: Hovorový kanál).



Graf 7.8: Hovorový kanál

U šumů vysavač, bílý šum a křižovatka se celková kvalita díky průchodu filtrem dokonce zvýšila, což je způsobeno tím, že filtr snížil intenzitu šumu v celkovém signálu a kvalitu samotné řeči ovlivnil minimálně, čímž vlastně zvýšil podobnost signálu k signálu referenčnímu, tedy čisté řeči bez šumu i filtru. Tyto závěry korespondují s výsledky poslechových testů, kde bylo patrné větší snížení kvality i signálu zfiltrovaných pásmovou zádrží.

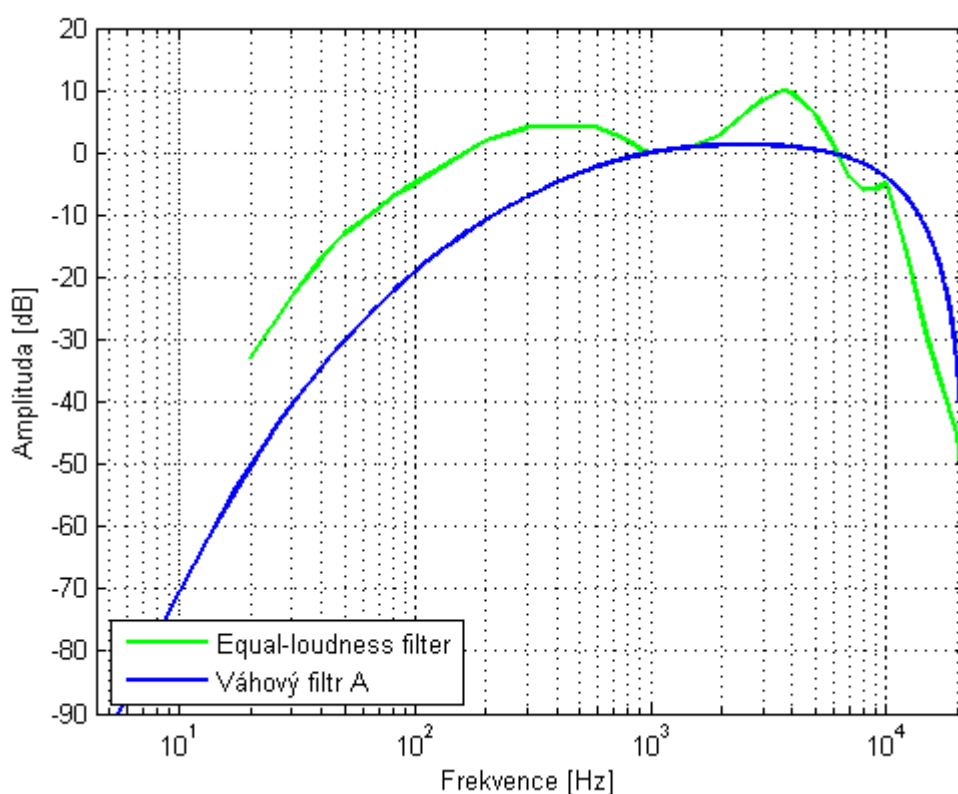
### **7.3 Srovnání poslechových testů a objektivních metod**

Chceme-li vyhodnotit kvalitu řečového, nebo všeobecně zvukového signálu, můžeme buď provést poslechové testy, nebo si vybrat nějakou z objektivních metod. Kterou variantu ale zvolit?

Největší výhodou poslechových testů je, že nám dávají nejvíce vypovídající výsledek. Na druhou stranu jejich provedení je velmi náročné v mnoha aspektech, které jsem zmínil v kapitole 3.1. Nejmarkantnější je jejich časová náročnost. Poslechových testů prováděných v této práci se zúčastnilo 20 osob s průměrnou dobou trvání 45 minut, což dává celkem 15 hodin. Oproti tomu vyhodnocení stejného počtu signálu pomocí metody PEMO-Q zabralo asi hodinu, což je výrazné snížení oproti poslechovým testům. Objektivních metod pro vyhodnocování kvality zvukových signálů je však velké množství a u samotných metod se dají nastavit různé parametry, což zapříčiní různé výsledky. V případě metody PEMO-Q jsem vyzkoušel 4 druhy měření a až porovnáním s výsledky z poslechových testů jsem zjistil, které měření se pro dané testovací signály hodí nejvíce. Proto je potřeba každou metodu pro objektivní vyhodnocování kvality zvukových signálů vyzkoušet na co nejrozmanitějších typech zvukových signálů a porovnat s poslechovými testy. Na výše položenou otázku bych tedy odpověděl: Kvůli velké časové úspoře zvolte některou z objektivních metod, ale to pouze v případě, že již byla ověřena a porovnána s poslechovými testy na stejném typu signálů, na které ji chcete použít vy.

## 8 Vyhodnocení hlasitosti řečových signálů

K vyhodnocení hlasitosti řečových signálů jsem použil váhový filtr A a equal-loudness filtr, které jsem aplikoval na testovací databázi. Oba tyto aparáty na úpravu hlasitosti fungují tak, že upravují amplitudu signálu v závislosti na jeho frekvenci. Způsob vytvoření a testy funkčnosti obou filtrů naleznete v mojí práci [3]. Na obrázku (Obr. 8.1: Váhový filtr a equal-loudness filtr) si můžete prohlédnout průběh obou filtrů.



Obr. 8.1: Váhový filtr a equal-loudness filtr

### 8.1 Čisté řečové vzorky

Nejprve jsem se zabýval vyhodnocením hlasitosti čistých nefiltrovaných řečových vzorků bez aditivního šumu. Hlasitost vzorků se pohybovala okolo -20dB. Nenašel jsem zde žádnou výraznou závislost hlasitosti na mluvčím. Po aplikaci váhového filtru se hlasitost průměrně zvýšila o 0,61dB, equal-loudness filtr hlasitost také zvyšoval a to průměrně o 2,64dB. V následující tabulce (Tab. 8.2: Vliv mluvčího na hlasitost) si můžete prohlédnout průměrné hodnoty hlasitostí pro jednotlivé mluvčí a korekce po aplikaci filtrů.

Tab. 8.2: Vliv mluvčího na hlasitost

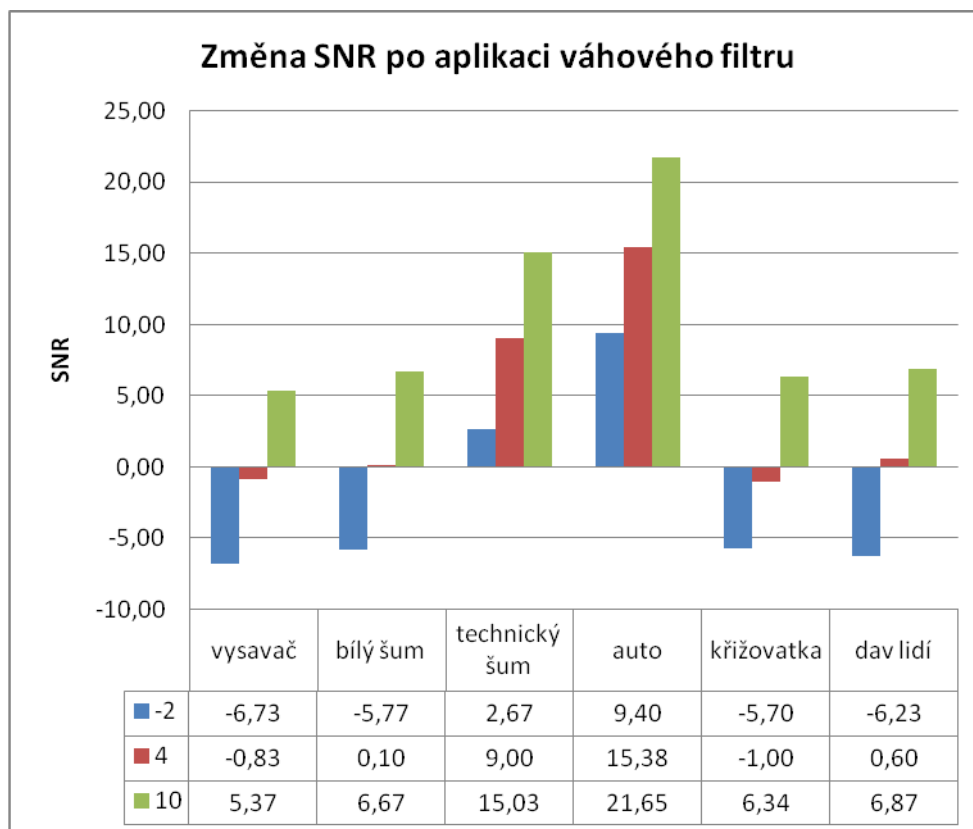
[dB]	Bez filtru	Váhový filtr	Equal-loudness filtr
muž1	-19,94	+0,50	+2,64
muž2	-21,00	+0,74	+2,20
žena1	-19,99	+0,55	+2,90
žena2	-19,47	+0,64	+2,82

## 8.2 Řečové vzorky s aditivním šumem

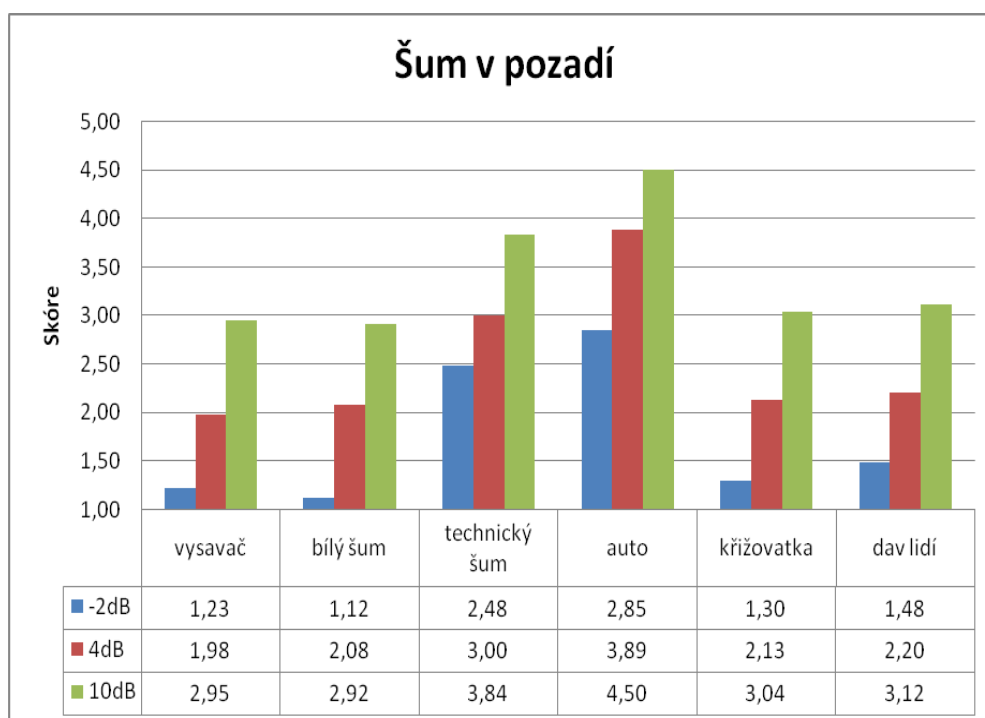
Dále jsem se zabýval hlasitostí řečových signálů s aditivním šumem. Šum se k řečovému signálu přičítá v závislosti na zvoleném SNR, které nám udává poměr řečového signálu a šumu v celkovém signálu. Aplikací váhového filtru nebo equal-loudness filtr se však tento poměr změní, protože filtry upravují řečový signál jinak než šum. Výpočet SNR po aplikaci váhového nebo equal-loudness filtru nám může vypovědět mnoho o skutečné hlasitosti šumu, čili o jeho vlivu na kvalitu celkového signálu. Výpočet SNR se provádí podle následujícího vzorce:

$$SNR = 10 \cdot \log \frac{\sigma_s^2}{\sigma_n^2},$$

kde  $\sigma_s^2$  je výkon užitečného signálu a  $\sigma_n^2$  je výkon šumu. Začneme vyhodnocením vlivu váhového filtru A na SNR. Výsledky jsou zobrazeny v následujícím grafu (Graf 8.3: Změna SNR po aplikaci váhového filtru). Graf zobrazuje, jak se změnila hodnota SNR po aplikaci váhového filtru, čili např. modré sloupky odpovídají hodnotám SNR -2dB před zfiltrováním a zobrazují hodnotu SNR po zfiltrování. Pro lepší porovnání s výsledky poslechových testů zde naleznete i graf zobrazující rušení šumu v pozadí signálů (Graf 8.4: Šum v pozadí).



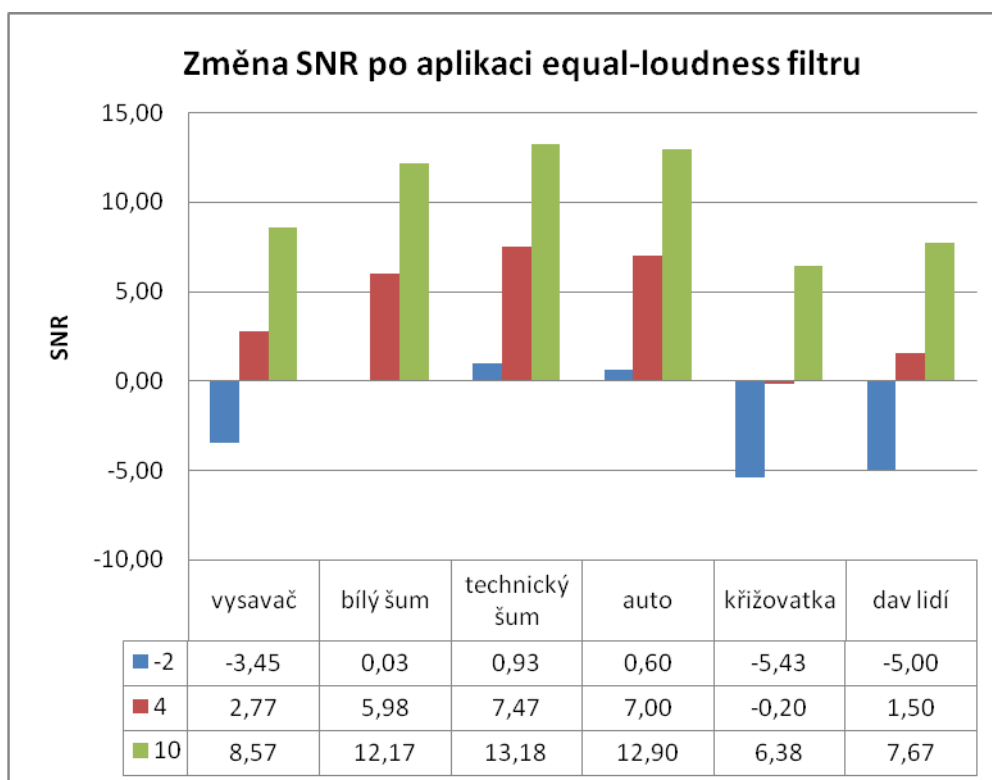
Graf 8.3: Změna SNR po aplikaci váhového filtru



Graf 8.4: Šum v pozadí

Je vidět výrazný nárůst SNR u technického šumu a ještě výraznější u šumu auto. Nárůst SNR ukazuje na snížení poměru šumu v celkovém signálu, čili na jeho menší hlasitost a tedy nižší rušení, což odpovídá i výsledkům poslechových testů. U ostatních šumů se SNR naopak snížilo, což signalizuje výraznější rušení těchto šumů.

Váhový filtr A je vlastně zjednodušená verze equal-loudness filtru, tudíž by měl equal-loudness filtr dávat lepší výsledky, ty jsou zobrazeny v grafu (Graf 8.5: Změna SNR po aplikaci váhového filtru).



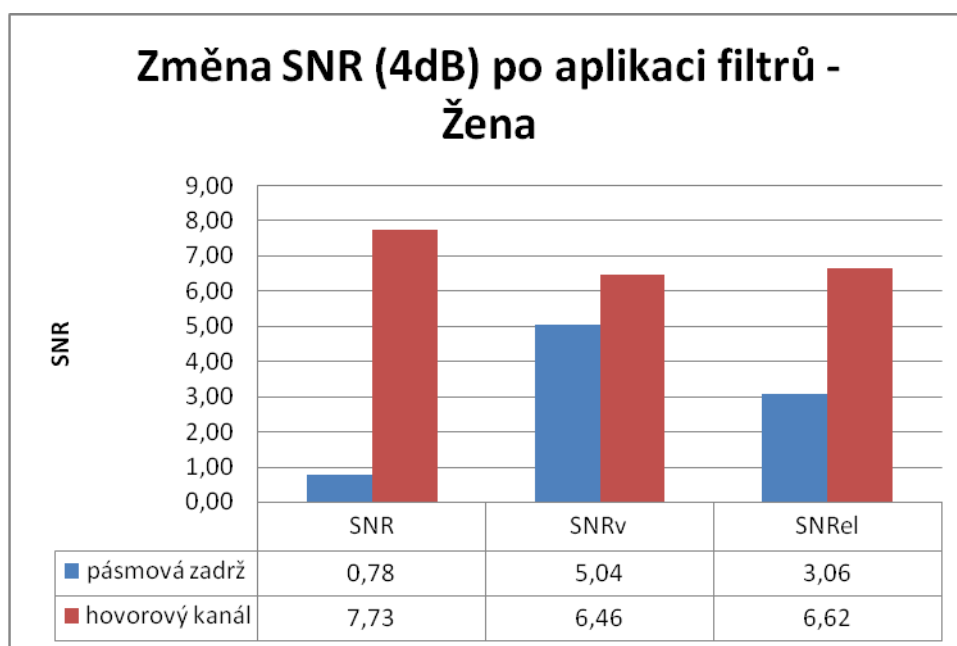
Graf 8.5: Změna SNR po aplikaci váhového filtru

Oproti výsledkům dosažených váhovým filtrem je zde nárůst SNR u stacionárních šumů vysavač a bílý šum, což je v rozporu s výsledky poslechových testů i s výsledky dosažených metodou PEMO-Q, kde stacionární šumy vykazovali větší rušení, čili horší kvalitu testovaných signálů. Jak už jsem zmínil, equal-loudness filtr by však měl dávat lepší výsledky než váhový filtr, což ale menší podobnost s výsledky poslechových testů nevyvrací. Hlasitost šumu je jen jedním z aspektů ovlivňujících jeho rušení, tedy i méně hlasitý šum může být více rušivý než šum hlasitější.

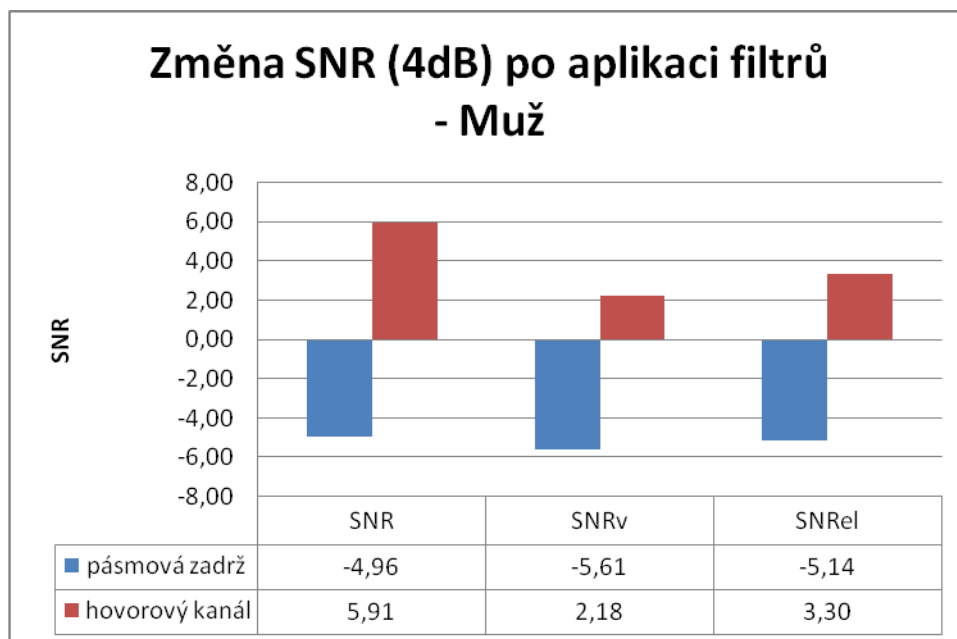


### 8.3 Zfiltrované řečové vzorky s aditivním šumem

Jak jsem zmiňoval v kapitole 5, řečové vzorky s aditivním šumem jsem filtroval dvěma filtry. Prvním z nich byla pásmová propust simulující telefonní kanál, druhým pásmová zadrž, která nepropouštěla frekvence v rozmezí 500Hz až 1kHz. Oba tyto filtry měly také vliv na poměr řečového signálu a šumu SNR. Jak tyto filtry ovlivňovaly SNR, jsem ověřil následujícím pokusem. Čtyři řečové vzorky, od každého ze čtyř mluvčích jeden, jsem postupně mísil se všemi šumy v poměru SNR rovnému 4dB. Na všechny vzniklé vzorky jsem aplikoval filtr pásmová propust a následně spočetl nové SNR (v grafu SNR), dále SNR po aplikaci váhového filtru A (v grafu SNRv) a SNR po aplikaci equal-loudness filtru (v grafu SNRel). Stejný postup jsem opakoval i u filtru pásmová zadrž. Výsledky pro řečníky označené žen1 a muž2 jsou zobrazeny v následujících grafech.



Graf 8.6: Změna SNR po aplikaci filtrů – žena



Graf 8.7: Změna SNR po aplikaci filtrů – muž

Je patrný poměrně velký rozdíl mezi oběma mluvčími, filtry ovlivňují každý hlas jinak. Filtr pásmová zadrž výrazně ovlivňoval kvalitu řečového signálu, můžeme si všimnout, že u mluvčího muž2 daleko výrazněji než u mluvčí žena1. Tím způsoboval snížení SNR z původní hodnoty 4dB na hodnoty zobrazené v grafu, čili klesala hlasitost vlastního řečového signálu. Filtr pásmová zadrž kvalitu celkového signálu snižoval, což koresponduje s výsledky poslechových testů i metody PEMO-Q. Oproti tomu filtr simulující hovorový kanál hodnotu SNR ve většině případů zvyšoval, což je způsobené tím, že snižoval míru šumu v celkovém signálu. Jedinou výjimkou byly vzorky namluvené mužem2, které dosahovaly ve všech testech horších výsledků.

## 9 Závěr

Diplomová práce se zabývá problematikou hodnocení kvality řečového signálu. Tato oblast je poměrně rozsáhlá, proto byly vybrány dvě, respektive tři metody jak kvalitu řečového signálu hodnotit. Prvním z těchto metod jsou poslechové testy jako zástupci subjektivních metod, z objektivních metod pro odhad vnímané kvality byla vybrána nová metoda PEMO-Q (je zde testována i metoda PEAQ, ze které metoda PEMO-Q vychází). V práci pak ještě naleznete část týkající se hlasitosti řečového signálu.

V úvodních kapitolách jsem se zabýval popisem řečového signálu, definicí jeho kvality a popisem vybraných metod. Praktická část práce začíná popisem tvorby databáze testovacích signálů. Tato databáze se skládá z 60 řečových signálů rušených šesti aditivními šumy (tři stacionární a tři nestacionární šumy) v různých poměrech SNR a zkreslených dvěma filtry. Na této databázi jsem pak testoval metody pro hodnocení kvality řečového signálu.

Nejprve jsem provedl poslechové testy a to podle doporučení ITU-T P.835. Testu se zúčastnilo 20 posluchačů s průměrnou dobou trvání jednoho testu 45 minut. Z testů vyplývá výrazná závislost rušení aditivního šumu na jeho poměrném zastoupení v celkovém signálu. Kvalita vlastního řečového signálu opět koresponduje s mírou zastoupení aditivního šumu, při nízké úrovni šumu je pak patrný vliv zfiltrování signálu, kvalitu výrazněji snižuje filtr pásmová zádrž. V porovnání stacionárních a nestacionárních šumu vítězí signály s nestacionárními šumy, které vykazují o něco lepší kvalitu. Výsledky poslechových testů jsem bral jako absolutní měřítko kvality testovaných signálů a využil jsem je k porovnání s objektivními metodami.

Databázi testovacích signálů jsem posléze vyhodnotil pomocí měření objektivní kvality řeči  $q_c$ , což je výstup metody PEAQ. Toto měření dosahovalo výsledků srovnatelných s výsledky poslechových testů. Lepších výsledků však bylo dosaženo pomocí metody PEMO-Q, kterou jsem testoval ve čtyřech variantách. Nejvíce vypovídajících výsledků na testovaných řečových signálech s aditivním šumem bylo dosaženo pomocí měření PSM a ve variantě s filtrem dolní propust. Toto měření jsem posléze využil k ověření vlivu zfiltrování signálu na jeho kvalitu. Filtr pásmová zádrž u všech signálů snižoval kvalitu, což koresponduje s poslechovými testy. Oproti tomu

filtr simulující hovorový kanál u třech šumů kvalitu dokonce zvyšoval, což si vysvětluji tím, že filtr snižoval míru šumu v signálu.

Shrnuli tedy funkčnost metody PEMO-Q, musím konstatovat, že metoda dosahovala při správné volbě měření a nastavení odpovídajících výsledků v porovnání s poslechovými testy.

V závěrečné části práce jsem se zabýval hlasitostí řečových signálů. Hlasitost je jeden z aspektů, který výrazně ovlivňuje kvalitu řečového signálu. Různé šумы, i když mají stejnou energii a jsou v signálu zastoupeny ve stejném poměru SNR, mohou být pro člověka různě hlasité a rušivé. Na druhou stranu i méně hlasitý zvuk může být více rušivý než zvuk hlasitější. Hlasitost ovlivňuje i průchod signálu nějakým filtrem, proto musíme brát v potaz, zda filtr ovlivňuje více řečový signál, nebo šum.

## Seznam použité literatury

- [1] Huber, R. – Kollmeier, B. *PEMO-Q—A New Method for Objective Audio Quality Assessment Using a Model of Auditory Perception*. IEEE transactions on audio, speech, and langure proccesing, vol. 14, no. 6, 2006.
- [2] Bech, S. – Zacharov, N. *Perceptual Audio Evaluation - Theory, Method and Application*, Wiley, ISBN: 978-0-470-86923-9, 2006.
- [3] Pavlík, Jan. Vyhodnocené hlasitosti zvukového signálu. Semestrální projekt. Fakulta mechatroniky, informatiky a mezioborových studií. Technická univerzita v Liberci. Liberec. 2012.
- [4] Rec. ITU -R BS.1116-1. *Methods for the subjective assesment od small impairments in audio system including multichannel sound system*. 1997.
- [5] Rec. ITU-T P.835. *Subjective test methodology for evaluating speech communication systems that includenoise suppression algorithm*. 2003.
- [6] Bezdíček, Martin. *Subjektivní a objektivní metody hodnocení řečového signálu*. Bakalářská práce. Fakulta elektrotechniky a komunikačních technologií. Vysoké učení technické v Brně. Brno. 2008.
- [7] Polák, Petr. *Metody odhadu odstupu signálu od šumu v řečovém signálu*. FEL ČVUT. Praha. 2001.
- [8] Zezula, Radek. *Objektivní a subjektivní metody vyhodnocování kvality vodoznačných audio signálů*. FEL Fakulta elektrotechniky a komunikačních technologií. Vysoké učení technické v Brně. Brno. 2008.
- [9] Bureš, Zbyněk. *Objektivní hodnocení kvality zvuku*. Katedra radioelektroniky. FEL ČVUT. Praha. 2006.
- [10] Bernat Petr a spol. *Akustika, vznik a šíření zvuku, frekvenční analýza a syntéza, sluchový vjem zvukového signálu*. Anatomie varhan [Online]. [cit. 2012-02-16] URL:[ [http://homen.vsb.cz/~ber30/texty/varhany/anatomie/pistaly\\_akustika.htm](http://homen.vsb.cz/~ber30/texty/varhany/anatomie/pistaly_akustika.htm)]

- [11] Jiříček Ondřej a spol. *Zvukoměr – Akustický analyzátor* [Online].  
URL:[[http://archiv.otevrena-veda.cz/users/Image/default/C1Kurzy/Fyzika/21\\_jiricek.pdf](http://archiv.otevrena-veda.cz/users/Image/default/C1Kurzy/Fyzika/21_jiricek.pdf)].
- [12] Heringová Blanka, Hora Petr. *Matlab – díl I: Práce s programem*. 1995.
- [13] Equal-loudness filter [Online]. [cit. 2012-02-17]. URL:[[http://replaygain.hydrogenaudio.org/proposal/equal\\_loudness.html](http://replaygain.hydrogenaudio.org/proposal/equal_loudness.html)].
- [14] A-weighting. [www.wikipedia.org](http://www.wikipedia.org) [Online]. [cit. 2012-02-08]. URL:[<http://en.wikipedia.org/wiki/A-weighting>].
- [15] Opršal, Petr. Zvýrazňování řeči na bázi transformace wavelet [Online]. [cit. 2012-02-08]. URL:[ <http://www.p007.webpark.cz/se/>], VŠB-TU, Ostrava, 2006.
- [16] HörTech gGmbH. *PEMO-Q manual*. Oldenburg, Germany, 2010.

## Příloha A – Instrukce k poslechovému testu

V následujícím experimentu uslyšíte 2 skupiny řečových vzorků. Každá skupina obsahuje 30 vět s různými druhy šumu v pozadí a s různými úrovněmi odstupu signálu od šumu. Tyto vzorky budete hodnotit.

Hodnocení bude probíhat následujícím způsobem:

Na každé stránce si třikrát poslechnete jeden řečový vzorek. Při prvním poslechu ohodnotíte kvalitu řeči podle následující tabulky (Tab. 1).

Tab. 1: Kvalita řečového signálu

Řečový signál je:	Skóre
Nezdeformovaný	5
Mírně zdeformovaný	4
Trochu zdeformovaný	3
Docela zdeformovaný	2
Velmi zdeformovaný	1

Při druhém poslechu budete hodnotit rušení pozadí a to podle následující tabulky (Tab. 2).

Tab. 2: Rušení pozadí

<b>Pozadí je:</b>	<b>Skóre</b>
Téměř nezatelné	5
Mírně znatelné	4
Znatelné, ale ne rušivé	3
Docela rušivé	2
Velmi rušivé	1

Při posledním poslechu ohodnotí celkovou kvalitu signálu podle následující tabulky (Tab. 3).

Tab. 3: Celková kvalita signálu

<b>Celková kvalita signálu je:</b>	<b>Skóre</b>
Vynikající	5
Dobrá	4
Průměrná	3
Nízká	2
Špatná	1

Zvuk přehrajete kliknutím na obrázek reproduktoru. Do záznamové tabulky budete zaznamenávat číselné hodnoty odpovídajícího skóre.



## Záznamová tabulka

### Část 1

Vzorek	Kvalita řečového signálu	Rušení pozadí	Celkové kvalita signálu
1			
2			
3			
4			
5			
6			
7			
8			
9			
10			
11			
12			
13			
14			
15			
16			
17			
18			
19			
20			
21			
22			
23			
24			
25			
26			
27			
28			
29			
30			

## Část 2

Vzorek	Kvalita řečového signálu	Rušení pozadí	Celkové kvalita signálu
1			
2			
3			
4			
5			
6			
7			
8			
9			
10			
11			
12			
13			
14			
15			
16			
17			
18			
19			
20			
21			
22			
23			
24			
25			
26			
27			
28			
29			
30			